

A STUDY ON ENHANCING PATIENT SATISFACTION AND SERVICE QUALITY THROUGH MACHINE LEARNING CLASSIFIERS AND BIG DATA ANALYTICS

A Thesis

Submitted towards the Requirements for the Award of Degree of

**DOCTOR OF PHILOSOPHY
IN
COMPUTER SCIENCE AND ENGINEERING**

Under the Faculty of Engineering and Technology

Submitted By

Mr. NAYAN SARKAR

(Enrollment No.: 220101165)

Under the Supervision of

Prof. (Dr.) BALVEER SINGH

P.K. UNIVERSITY



2026

**P.K. University
Nh- 27, Vill. Thanra (P.O. Dinara)
Shivpuri, M.P. India - 473665**



P.K.UNIVERSITY
SHIVPURI (M.P.)

University Established Under section 2F of UGC ACT 1956 Vide M.P. Government Act No17 of 2015

CERTIFICATE OF THE SUPERVISOR

This is to certify that the work entitled “**A STUDY ON ENHANCING PATIENT SATISFACTION AND SERVICE QUALITY THROUGH MACHINE LEARNING CLASSIFIERS AND BIG DATA ANALYTICS**” is a piece of research work done by **Mr. NAYAN SARKAR (Enrolment No.: 220101165)** Under my Guidance and Supervision for the degree of Doctor of Philosophy in Computer Science & Engineering at P.K. University Shivpuri (M.P) India.

I certify that the candidate has put an attendance of more than 240 days with me. To the best of my knowledge and belief the thesis:

I – Embodies the work of the candidate himself

II – Has duly been completed.

III –Fulfil the requirement of the ordinance relating to the Ph.D. degree of the University.

Date:/...../.....

Signature of the Supervisor



P.K.UNIVERSITY
SHIVPURI (M.P.)

University Established Under section 2F of UGC ACT 1956 Vide M.P. Government Act No17 of 2015

DECLARATION BY THE CANDIDATE

I declare that the thesis entitled “**A STUDY ON ENHANCING PATIENT SATISFACTION AND SERVICE QUALITY THROUGH MACHINE LEARNING CLASSIFIERS AND BIG DATA ANALYTICS**” is my own work conducted under the supervision of **Prof. (Dr.) BALVEER SINGH** Department of Computer Science & Engineering, P.K. University, Shivpuri (MP), India.

Approved by Research Degree Committee. I have put more than 240 days of attendance with supervisor at the center.

I further declare that to the best of my knowledge the thesis does not contain my part of any work has been submitted for the award of any degree either in this University or in any other University Without proper citation.

Date:/...../.....

Signature of the Candidate

Place : P.K. University



P.K.UNIVERSITY
SHIVPURI (M.P.)

University Established Under section 2F of UGC ACT 1956 Vide M.P. Government Act No17 of 2015

FORWARDING LETTER OF HEAD OF INSTITUTION

The Ph.D. thesis entitled “A STUDY ON ENHANCING PATIENT SATISFACTION AND SERVICE QUALITY THROUGH MACHINE LEARNING CLASSIFIERS AND BIG DATA ANALYTICS” Submitted by **Mr. NAYAN SARKAR (Enrolment No.: 220101165)** is forwarded to the university in seven copies. The candidate has paid the necessary fees and there are No-Dues outstanding against him.

Name Seal

Date:/...../.....

Place: P.K. University

(Signature of Head of Institution where the
Candidate was Registered for Ph.D degree)

Signature of the Supervisor

Date :.....

Place:- P.K. University



P.K.UNIVERSITY
SHIVPURI (M.P.)

University Established Under section 2F of UGC ACT 1956 Vide M.P. Government Act No17 of 2015

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: **“A STUDY ON ENHANCING PATIENT SATISFACTION AND SERVICE QUALITY THROUGH MACHINE LEARNING CLASSIFIERS AND BIG DATA ANALYTICS”**

Candidate's Name: **Mr. NAYAN SARKAR (Enrolment No.: 220101165)**

COPY RIGHT TRANSFER

The undersigned hereby assigns to the P.K. University, Shivpuri (M.P.) all copyrights that exist in and for the above thesis submitted for the award of the Ph.D. degree.

Date :.../..../.....

Name of the Scholar

Note : However, the author may reproduce/publish or authorize others to reproduce, material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the University's copyright notice are indicated.

ACKNOWLEDGEMENT

The Acknowledgement of a research endeavor is an opportunity to express appreciation for the joint effort of those who have been sources of motivation and support and who have contributed invaluable knowledge toward achieving success in the pursuit of knowledge. With utmost humility and pride, I reflect on the inspiring individuals who have propelled my work forward, and I am filled with immense pleasure, gratitude and thanks for their unwavering dedication.

First and foremost, I would like to thank almighty God for perseverance, courage and strength bestowed on me to undertake this project.

I would like to extend my deepest appreciation and gratitude to my esteemed Supervisor, **Prof. (Dr.) Balveer Singh**, P.K. University Shivpuri (M.P.) without whom this work should not have come to fruition. His exceptional guidance, unwavering support kind co-operation, and provision of a conducive work environment have been invaluable beyond measure. I remain indebted to him for instilling in me a relentless pursuit of excellence, a strong sense of honesty and respect for ethical principles that govern our profession.

Furthermore, I extend my sincere thanks to **Shri Jagdeesh Prasad Sharma** Chancellor, P.K. University Shivpuri (M.P.) for providing the necessary infrastructure and facilities that were essential during my research work. My profound sense of gratitude, faith and awe also goes to **Prof. (Dr.) Yogesh Chandra Dubey** Vice Chancellor, P.K. University Shivpuri (M.P.) for his generous and valuable support throughout my research. I express my gratitude to **Prof. (Dr.) Mukesh Chansoriya** Pro Vice-Chancellor, P.K. University Shivpuri (M.P.) for his generous and valuable support throughout my research.

I take this pleasant opportunity to express my deep sense of gratitude and reverences to respectable Hon'ble **Dr. Jitendra Kumar Mishra** Director; P.K. University Shivpuri (MP) for his constant encouragement and providing me with the required facilities that enabled me to complete my project work.

I also extend my deepest appreciation **Prof. Jitendra Kumar Malik**, Registrar of P.K. University Shivpuri (M.P.) has inspired me with his encouraging guidance and support right from the beginning of this work despite his busy schedule. I am short of words to thank him for his affectionate behavior and patience throughout the duration of the research work.

I express my gratitude to **Dr. Aiman Fatima**, Dean of Academics of P.K. University Shivpuri M.P. I owe a great deal of application and gratitude to the entire faculty members and all non-teaching staff of the department without their support it would have been difficult to shape my research work.

I would also like to extend my gratitude and deep appreciation to respected **Dr. Ashish Vishwakarma**, Dean Research of P.K. University, M.P.

I would also like to thank **Dr. Monika Tripathi**, P.K. University, M.P. and **Mr. Ashish Pratap Singh** and **Mr. Neeraj Gaur** for their valuable support and guidance.

I also express my gratitude to **Miss Nisha Yadav**, Librarian of P.K. University along with Mr. Anand Kumar, Assistant Librarian and other staff of P.K. University. I am also thankful to **Mr. Pankaj Singh**, **Mr. Alok Khare** and **Mr. Purushottam Yadav**, I. T. Cell of P.K. University.

Date:....../...../.....

NAYAN SARKAR

Place: P.K. University



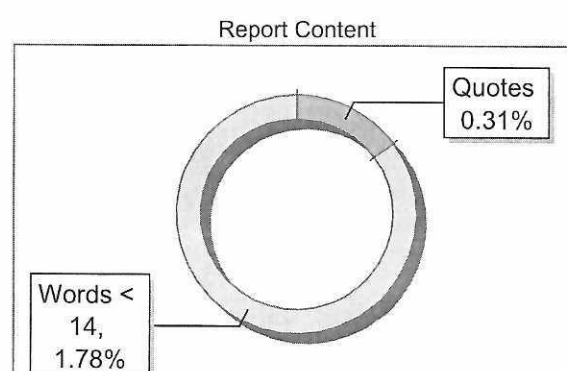
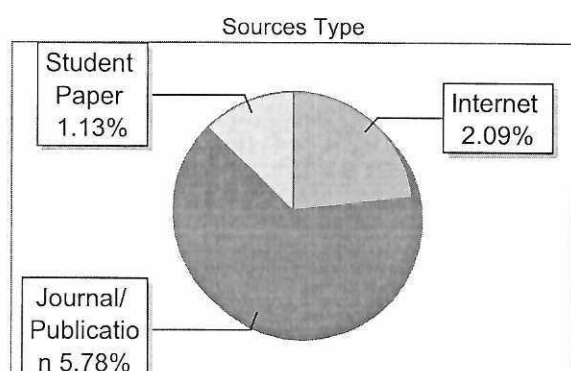
The Report is Generated by DrillBit Plagiarism Detection Software

Submission Information

Author Name NAYAN SARKAR
Title A STUDY ON ENHANCING PATIENT SATISFACTION AND SERVICE QUALITY THROUGH MACHINE LEARNING CLASSIFIERS AND BIG DATA ANALYTICS
Paper/Submission ID 5436771
Submitted by library.pku@gmail.com
Submission Date 2026-04-07 14:09:59
Total Pages, Total Words 151, 37373
Document type Thesis Chapter

Result Information

Similarity 9 %



Exclude Information

Quotes	Not Excluded
References/Bibliography	Not Excluded
Source: Excluded < 14 Words	Not Excluded
Excluded Source	0 %
Excluded Phrases	Not Excluded

Database Selection

Language	English
Student Papers	Yes
Journals & publishers	Yes
Internet or Web	Yes
Institution Repository	Yes

A Unique QR Code use to View Download Share Pdf File

LIBRARIAN
P.K. University
Shivpuri (M.P.)



A Study on Enhancing Patient Satisfaction and Service Quality through Machine Learning Classifiers and Big Data Analytics

Abstract

In order to tackle the complicated problems of improving service quality and patient happiness, the healthcare sector is increasingly using technology. The purpose of this research is to find out how healthcare providers and patients might benefit from using big data analytics and machine learning classifiers. In order to determine what factors most influence patients' levels of satisfaction, this study's primary objective is to build prediction models that can sift through massive datasets such as survey responses, medical records, and operational indicators. One of the primary goals of this project is to use patient satisfaction data to develop a novel feature selection technique for home health care services. The first step in using sentiment analysis to patient service satisfaction is to provide a systematic overview of the literature. This overview centers on health-related opinion polls and review comments, as well as features linked to patient satisfaction and feedback on medication recommendations and reviews. The second contribution of this project is a new SCSP ensemble model for analyzing health care opinion systems among patients. By using a machine learning strategy to forecast superior models in data analysis, the classification models categorize patients' emotions as positive, negative, or neutral. By using ensemble methods to the model-classified views, we may examine the model's health care suggestion in light of the sentiment polarity. Based on the tweets, we have chosen this issue in order to provide the healthcare business a solid information system. After collecting data on patient satisfaction, this study presents a new feature selection approach for determining which aspects of home health care services are most important. In this, we looked at the many parts that make up health care service superiority evaluations using various metrics. We use a variety of machine learning algorithms to try to predict what factors have the most impact on patients' happiness with their treatment. It is well-known that evaluating healthcare services based on the patient's experience is crucial. The next section of this study proposal details a new approach to sentiment analysis that uses symptoms to assess patients' levels of medication satisfaction via the suggestion technique. The next step is to compare the suggested methods' performance to that of the current systems, and finally, to provide the reasons for the suggested methods. When applied to healthcare industry social media evaluations,

our suggested approach achieves an AUC accuracy of 95% and an overall accuracy of 93%. Based on the number of favorable evaluations, it gives a recommendation % for the healthcare industry. In addition to presenting the findings of the suggested model's analysis, this study offers model verification via the validation of the models used. As a final section, the new aspects and contributions make up the proposed work's general conclusion. As an added bonus, this chapter lays out the suggested methods for future study as well as outstanding problems in the health care arena. The results show that operational efficiency, service quality, and patient happiness are all much improved when big data analytics and machine learning classifiers are used together. In order to guarantee sustainable and excellent healthcare delivery in an increasingly digital age, the report recommends that healthcare organizations incorporate these technologies into a patient-centric strategy.

Keywords : Patient, Healthcare, Machine Learning, Big Data, Model

LIST OF CONTENTS

Sr. No.	Chapter Name	Page No.
1.	Introduction	1-45
2.	Review of Literature	46-79
3.	SCSP Ensemble Model and Feature Selection for Predicting Patient Health Care Opinions and Home Health Care Service	80-126
4.	Sentiment Analysis and Performance Comparison for Evaluating Patient Opinions on Drug Satisfaction	127-145
5.	Findings, Conclusion, Suggestions and Future Scope of The Study	146-151
–	References	152-161

LIST OF FIGURES

Figure No.	Figure Name	Page No.
1.1	Service quality – expectations and perceptions	28
1.2	Classification of big data	31
3.1	Class Distribution on Reviews	83
3.2	Assessment centered on Healthcare Facilities	84
3.3	Design of the SCSP Model Being Proposed	85
3.4	Top 20 terms used in hospital assessments prior to the elimination of stop words	89
3.5	The top twenty terms removed from the hospital review	89
3.6	Prior to removing stop words, the top 20-bigrams in hospital reviews	90
3.7	Twenty most important terms removed from hospital evaluations	90
3.8	An ensemble model for SCSP	93
3.9	Comparison of Model before ensemble	98
3.10	Comparison of Model after ensemble	98
3.11	ROC curve	100
3.12	Prediction Metric Score	100
3.13	Proposed Framework	104
3.14	Scatter Plot	113
3.15	Correlation Matrix	114
3.16	Ranking based on features	115
3.17	Classification report of Logistic regression	117

3.18	Classification report of Random Forest	118
3.19	Classification report of Gradient Boost	119
3.20	Classification report of Ada Boost	120
3.21	Feature Importance based on model	121
3.22	Model Comparison	122
3.23	ROC	123
3.24	Features based on HHCS	126
4.1	Proposed Methodology	128
4.2	User Ratings (1-10)	130
4.3	User Ratings (1/0)	131
4.4	Performance Metrics Comparison for Negative and Positive Classes on LG Boost	133
4.5	Performance Metrics Comparison for Negative and Positive Classes of XG Boost	134
4.6	Performance Metrics Comparison for Negative and Positive Classes of Cat Boost	135
4.7	Accuracy	136
4.8	Top drugs available per conditions in patients	137
4.9	Prediction Metric Score	138
4.10	Comparison - ROC-AUC curve	139
4.11	Top 10 drugs	139
4.12	Graph showing Comparison on healthcare service recommendation based on accuracy	144

LIST OF TABLES

Table No.	Table Name	Page No.
3.1	Accuracy of ML Classifiers on a 5-Class Set	93
3.2	Accuracy of ML Classifiers in a Three-Class Set	93
3.3	Prior to ensemble, model classification	96
3.4	Model classification after Ensemble	96
3.5	Classifier Model scores based on reviews	99
3.6	Hospital recommendation	101
3.7	Total number of respondents	106
3.8	Feature description before selecting the features	107
3.9	Feature description after exclusion of features	108
3.10	Features based on sentiments	109
3.11	Feature score for scatter and correlation matrix	112
3.12	Ranking based on features	115
3.13	Logistic regression classification report	116
3.14	Random Forest	117
3.15	Gradient boost	118
3.16	Ada Boost	119
3.17	Model performance using ROC	123
4.1	User Ratings (1-10 Scale)	130
4.2	No. of user ratings after conversion	130

4.3	Classification report on LG Boost	133
4.4	Classification report of XG Boost	134
4.5	Classification report of Cat Boost	135
4.6	Prediction metric scores for sentiment boosting classifier models	138
4.7	Comparison on healthcare service recommendation based on accuracy	143
4.8	Prediction metric scores for sentiment boosting classifier models	145

LIST OF ABBREVIATIONS

Abbreviations	Full Form
EHRs	Electronic health records
ASL	Analytic Server Language
TFPC	Traffic Prediction Model Called
HQR	Hospital Queuing Recommendation
KNN	K-Nearest Neighbors
SVM	Support Vector Machine
DT	Decision Tree
ASMs	Attribute Selection Measures
NB	Naïve Bayes
SVM	Support Vector Machine
NN	Neural Networks
LR	Logistic Regression
ICA	Independent Component Analysis
PCA	Principal Component Analysis
CRM	Customer Relationship Management
ML	Machine Learning
BDA	Big Data Analytics
LightGBM	Light Gradient Boosting Machine
BoW	Bag-of-Words

TF-IDF	Term Frequency-Inverse Document Frequency
RF	Random Forest
HEADR	Hybrid Enhanced Adaptive Data Rate
LDA	Latent Dirichlet Allocation
NLP	Natural Language Processing
PPIs	Proton Pump Inhibitors
(RCTs)	Randomized Clinical Trials
UMLS	Unified Medical Language System
PLS	Partial Least Squares Route Modeling
TP	True Positives
TN	True Negatives
FP	False Positives
FN	False Negatives

INTRODUCTION

1.1 Overview

Satisfaction of patients and high-quality service have become paramount in the ever-evolving healthcare sector. Healthcare professionals have recognized the incorporation of advanced technology as a potentially effective strategy to achieve their goal of delivering optimum care. There has been a lot of buzz about machine learning classifiers and big data analytics because of the revolutionary changes they might make to healthcare.

With the twin goal of improving medical operations and the patient experience, technological innovations in healthcare are driving a major paradigm change. The importance of measuring patient satisfaction—which includes treatment, communication, and outcomes—as a key indicator of healthcare quality is being recognized more and more. The effectiveness and efficiency of healthcare provision may be gauged by the degree to which patients are satisfied with the services they get. Identifying and addressing the factors that impact patient satisfaction and service quality is crucial for healthcare practitioners striving to provide patient-centered care.[1]

Machine learning, one of artificial intelligence's numerous subfields, has shown adept at extracting useful patterns from large, complicated datasets. By evaluating large volumes of patient data using machine learning classifiers, healthcare providers are able to make more precise predictions and provide individualized therapies. The crux of big data analytics, on the other hand, is mining massive databases for useful patterns. When used together, machine learning classifiers and big data analytics have the potential to optimize healthcare operations, improve clinical results, and—most importantly—make patients happier.

The capacity to identify and fix relevant elements is necessary for patient satisfaction, which is a challenging problem. In order to spot patterns in patient data, healthcare providers could use machine learning classifiers and patterns associated

with patient dissatisfaction. These classifiers can sift through mountains of data, including operational records, EHRs, and patient feedback, to unearth subtle connections that might otherwise go undetected by more traditional methods of data analysis. By understanding the many factors that influence patient happiness, healthcare organizations may tailor treatments and implement targeted improvements.

Machine learning classifiers have many uses in healthcare, including predicting patient outcomes and resource needs, as well as detecting sources of discontent. As a cornerstone of machine learning, predictive analytics enables healthcare professionals to anticipate patient needs and allocate resources more efficiently. Patient outcomes are improved and service quality is elevated when a proactive strategy is used. By using machine learning algorithms, healthcare practitioners may tailor care methods based on patient preferences, treatment responses, and potential problems. This leads to a more satisfying and patient-centered experience for patients.

Together with machine learning, the ever-increasing volume of healthcare data may be better understood and managed with the help of big data analytics. Big data analytics' scalability and adaptability allow healthcare firms to easily manage, study, and draw valuable insights from a diverse range of data sources, including patient demographics, diagnostic images, medical records, and social determinants of health, among many others. By thoroughly analyzing patient profiles, this all-encompassing method for a more sophisticated evaluation of service quality and patient happiness. [2]

Combining big data analytics with machine learning classifiers also allows for the construction of prediction models of patient satisfaction. Healthcare providers may construct multi-factor prediction models that assess the probability of patient satisfaction by examining historical data for patterns linked with good patient experiences. Healthcare businesses may benefit greatly from these approaches in their quest for proactive problem-solving, service-simplification, and patient happiness.

Sentiment analysis of patient comments might benefit greatly from the combination of predictive modeling, machine learning classifiers, and big data analytics. Analyzing textual data from social media, surveys, and patient reviews

allows healthcare providers to measure patient mood and find areas for improvement. Natural language processing techniques enable the extraction of valuable information from texts that lack a conventional organizational structure. This area of study falls within the machinery learning umbrella. The patient's emotional condition and life experiences may then be better understood. Insights from this data might be invaluable in crafting problem-specific actions that enhance service quality.

The benefits of big data analytics and machine learning go well beyond the domain of individual patient interactions and influence the whole healthcare ecosystem. Improving the health outcomes of a certain population is the goal of population health management, which may be greatly advanced with the help of these technological advancements.

Machine learning algorithms can foresee illness patterns, identify at-risk populations, and devise preventive strategies by evaluating data at the population level. Big data analytics strengthens population health programs and allows for the development of focused interventions to reduce health inequalities and improve overall well-being by providing a holistic view of the socioeconomic determinants of health.

There are still certain obstacles to using big data analytics and machine learning classifiers in healthcare settings, despite the fact that they hold great potential for improved service and happier patients. Privacy, data security, and ethical considerations must always take precedence when working with sensitive patient information. Strict security procedures and adherence to data protection requirements are necessary to guarantee the seamless integration of these technologies. Furthermore, doctors and other healthcare workers must be able to comprehend and implement the results of machine learning and big data analytics.

The merging of big data analytics with machine learning classifiers will revolutionize healthcare delivery and improve patient satisfaction. With the help of these innovations, healthcare providers may take a more proactive and individualized approach to patient treatment, moving beyond the conventional wisdom of the past.

By responding to particular concerns, improving service delivery based on patient input, and anticipating patient outcomes, healthcare providers might

potentially raise the bar for patient satisfaction. In light of the ever-evolving healthcare industry, one potential approach to achieving patient-centered, data-driven healthcare is to integrate machine learning classifiers with big data analytics.

1.2 Concept of Service Quality

1.2.1 Defining Service

In contrast, services are not warehousable but rather created in the process of consumption; so, there is no room for a commercial intermediary in the sale of these services; and finally, only direct sales are feasible. The marketing of services such as travel agents and insurance sometimes makes use of agent middlemen.

The majority of healthcare services are intangible, such as the knowledge and experience of the doctors, the quality of the hospital itself, the tools and machinery used, the friendliness of the staff, the orderliness of the patient rooms and bathrooms, etc. However, service products can also consist of tangible items, such as eyeglasses, a prosthetic device, prescription medications, or laboratory results.[3]

In contrast to physical goods, intangible services cannot be kept and do not provide ownership. According to Zeithaml et al. (2006), the services offered are considered intangible due to the limitations in marketing them as future performance. Most of the research presupposes that intangibility influences consumers' actions, especially their purchasing habits.

Every market's bicycle represents service, with its two wheels representing giving and taking, or meeting someone's need and receiving money in exchange. Because it is intangible, service can only be felt, thought about, or lost in memory; it cannot be physically experienced. Many of the services that businesses provide are intangible, meaning that customers cannot physically experience them.

1.2.2 Defining Quality

For the most part, when people purchase curd in bulk, they do so with the expectation that it would stay fresh for as long as the expiration date on the container says.

A client will know the curd is of low quality if their expectations aren't satisfied before the expiration date has passed and the curd has spoiled. Customers would stop trusting curd vendors to always provide fresh curd if this keeps happening.

Customers may be hesitant to make reliable forecasts due to uncertainties over the curd's freshness and consistency. Companies and customers alike understand that high-quality goods and services are essential, and that these factors may help differentiate one another's offerings.

Typically, the most reliable measure of quality is the degree to which customers are satisfied. Quality is a measure of how great something is when compared to others of the same sort. In business, engineering, and manufacturing, the practical definition of quality is an object's superiority or absence of inferiority. Additionally, a product that not only accomplishes its purpose but also fulfills customer expectations is also noted.

People tend to be skeptical about novel ideas and approaches. Reliability is something that everyone values. Life on Earth is complicated and full of surprises. All of a business's quality improvement efforts should be systematic, time-bound, and continuous.

It needs to be ingrained in your entire way of life and become automatic. Managers in every industry, from accountancy to manufacturing, have an obligation to continuously enhance the quality of their operations. If you want people to care about making quality better, you need to get down to brass tacks first. Help them understand what quality is and why quality management is important.[4]

If you want to know how to communicate and what to do to enhance the quality of your work, go no further than the 14 stages laid out by quality guru Philip Crosby for the process of term improvement:

1. **Management Quality:** For the organization's quality to improve, the management must be devoted and committed. For the correct mindset to develop toward a zero-defect product or service line, it is essential that the company's commitment to improvement be transparent and easy to understand for all personnel.

2. **Quality Improvement Team:** Members of the search team who can demonstrate a dedication to quality improvement must be assembled by the department heads. Designating one of them as the team's leader achieves excellent results.
3. **Quality Measurements:** Prior to developing a strategy to enhance product or service quality, it is crucial to assess the existing level of quality. At this point, it has been determined where improvements are possible and where room for improvement exists.
4. **Cost of Quality Evaluation:** Can you tell me how much quality costs and how much non-conformance costs? It is possible to convince everyone in an organization of the need of a quality improvement system by answering all of these questions.
5. **Quality Awareness:** Employees must now be made more aware of the need of quality improvement. Training and communication tools like posters, pamphlets, videos, etc., may help with this. The company's attempts to increase quality will be successful if people prioritize quality.
6. **Corrective Action:** It is necessary to identify a company's quality concerns. Making corrections to the identified issues is the next stage. Using causal analysis methodologies, the organization must guarantee that these problems do not emerge and continue to exist.
7. **Establish an Ad Hoc Committee for Zero Defects Program:** You should establish a committee consisting of three or four team members to ensure that your goods and services follow the zero-defects philosophy.
8. **Supervisor Training:** Verify that all of the company's supervisors are capable of doing their tasks in a way that ensures the quality of the products and services they provide. Zero defect status is more likely to be attained if the company's supervisors get good training.
9. **Zero Defect Day:** A zero faults day is a great opportunity to let everyone in the firm know how much we've improved our quality control measures.

10. **Goal setting:** Everyone in charge of the organization should be asked to set personal goals when the changes are made. Goals should be specific and quantifiable and often set for 30, 60, and 90 days. It would be great if managers and workers could work together to set goals. More dedication to reaching the goal of zero faults may be attained in this way.
11. **Error Cause Removal:** Any endeavor to enhance quality must prioritize the elimination of mistake sources if it is to achieve its intended outcome. It is important to encourage employees to approach corporate management if they encounter any problems with achieving their objectives.
12. **Recognition:** Employees that surpass expectations and achieve their objectives should be recognized with awards. It is recommended that no monetary prizes or awards be given. Employees really value it when their performance is acknowledged.
13. **Quality Councils:** Gather quality experts and team leaders on a regular basis to discuss ways to achieve quality improvement objectives, brainstorm potential solutions, and settle on a course of action to enhance the quality program that is already in place.
14. **Do it over Again:** The program is infinitely long due to repetition. The business must repeat the process indefinitely if it wants to raise the bar on product and service excellence.

1.2.3 Defining Service Quality

The value that a service provides to customers is dictated by how good it is. Companies demonstrate high service quality when they fulfill or beyond customers' expectations. In this context, "expectation" is how well the service provider meets the client's needs throughout service delivery, and "perception" means how the provider evaluates that performance. We were able to raise the bar on service quality by working together. To start, new concepts on service quality have emerged in response to the growing significance of service industries to world economies. Secondly, as a result of the deregulation of industries and professional services, there is a significant

need for innovations in promoting services. Thirdly, service quality is considered a critical business component in the industrial and information technology sectors.[5]

Customers' perceptions of a service provider's and those services' relative excellence or lack thereof are often thought of as indicative of service quality, which is often associated with customers' overall feelings about the firm.

Dimensions of Service Quality

There are five main criteria used to evaluate service quality:

- **Tangibility:** The service provider's physical location, tools, and personnel's demeanor are all part of this.
- **Reliability:** The reliability and accuracy with which the service provider can carry out the promised service.
- **Empathy:** It is the considerate, personalized attention that each client receives from the service provider.
- **Assurance:** Guarantees everything from the honesty and reliability of the company's employees to the trust and dedication of its clientele.
- **Responsiveness:** It is the eagerness of the company's employees to assist clients and provide them with prompt service.

Boulding, Kalra, Staelin, and Zeithaml's (1993) dynamic process model of service quality provided empirical evidence that service quality influences customers' desired behaviors in a favorable way. It is well acknowledged that in service contexts, consumer purchase intentions are largely shaped by service quality and customer happiness.

Expectations also have a big role in how people view the quality of the service they get. Negative word-of-mouth may spread quickly when customers' expectations aren't fulfilled. Customers' increasing standard of living is elevating the service sector to a more central position. Suppliers that prioritize service quality are always on the go to meet the demands of their clients by supplying and delivering produced things

with additional value. They need to keep providing excellent service and value to their clients if they want to remain in the market.

Improving service quality requires attention to both actual and intangible assets. The capacity to consistently deliver on promises of high quality is what we mean when we talk about reliability as a differentiating feature. Reliability is a measure of a service provider's ability to consistently and correctly provide their services. There is a positive and substantial relationship between customer happiness and the five dimensions of service quality: tangibles, responsiveness, assurance, and empathy. When it comes to this matter, the order of priority is as follows: tangibles, responsiveness, reliability, and empathy.

Any business that wants to succeed in today's global economy must prioritize quality. The quality of a service is distinct from the service itself. Service is more like a conversation, while quality is more like a need.[6]

1.3 Service Quality in Health Services

The supplier is essentially free to do anything he wants when it comes to decision-making. Worse still, healthcare production cannot be reduced to a simple unit count. In healthcare, the idea of output is not well-defined. Because of this paradigm, doctors and consumers alike often equate the cost of care with its quality, despite the fact that high-tech medical treatment is often very pricey. It is important to strive for providing affordable care of a fair quality.

There are three fundamental aspects of a human trajectory, one of which is a long and healthy life. At 129th in the world, India's Human Development Index is rather low. Government expenditure has to increase immediately for the public health-care system. No country, including India, can afford to put its public health care system in the hands of a private entity, such as a hospital. Over the last few years, we have seen a shift in the expectations of the public for improved healthcare facilities, driven by rising earnings. Hospitals, in order to attract and keep patients in this age of intense competition, have made improvements to the quality of their services a top priority (Alhashem et al., 2011; Arasli et al., 2008). Hospitals must raise the bar on quality if they are to live up to their patients' expectations. Consequently, quality

management has evolved into a crucial strategy for competing with competitors and ensuring the long-term viability of enterprises.[7]

1.3.1 Application of Service Quality to Healthcare Sector

Hospitals and other healthcare institutions' expansion in terms of revenue, market share, and public perception is strongly influenced by the quality of the services they provide. While many hospitals provide similar services, what sets them apart is the quality of those services. In addition, patients have higher expectations now that they are aware of the available alternatives and the rising quality of service. Because the hospital's bottom line is directly tied to the happiness and loyalty of its patients, it is crucial to measure their perceptions of the quality of the hospital's services.

Healthcare quality in hospitals may be categorized into two parts: functional or service quality and clinical/technical quality. Technical service quality evaluates the precision of medical diagnoses, the frequency and severity of invasive and non-invasive events, and the competence and experience of medical professionals. It places an emphasis on clinical scorecards, protocol adherence, and the tracking of clinical outcomes. Additionally treatment delivery and clinical outcomes may be enhanced via clinical audits by methodically reviewing treatment according to established procedures and using change management strategies. It is important to conduct clinical audits to verify that medical and paramedical professionals adhere to care protocols as outlined in their procedures. The audit plan of the hospital should include several clinical audits, such as those pertaining to patient safety, unique to the institution's specialty, and clinical governance.

The major focus of functional services is on key performance indicators and the efficacy of non-clinical service delivery. Technical services are more important to them than functional ones, and they have less knowledge about how to evaluate the diagnostic and treatment processes in terms of care quality. Technical quality does not reflect patients' views of health care quality since people do not comprehend clinical practice. The service quality approach has been the de jure standard for evaluating healthcare services; this approach focuses on the measurable, observable, and controllable components of quality that are important to patients. Quality of

healthcare services has become more important due to recent shifts in emerging nations' health policies. Patient satisfaction and assessments of service quality have a significant impact on healthcare service demand. Healthy rivalry in the health care market drives up standards of competence, productivity, and patient outcomes. With more healthcare providers offering similar services and people being aware of this, patients are more inclined to choose a hospital with good ratings or satisfaction scores from previous patients. In order to attract and retain patients, hospitals are focusing on improving treatment quality and outcomes. You may make the healthcare system more responsive by enhancing the quality of treatment and the patient's experience or enjoyment while obtaining services. In healthcare, the value of patient feedback and suggestions in gauging treatment quality has been a topic of contention.[8]

The primary goal of public hospitals is to serve the general population by offering affordable healthcare. Because of this, they may stop caring about the quality of their services and start focusing only on getting them to the people who need them. Still, public hospitals are starting to pay more attention to how their services are measured and evaluated; yet, there's room for improvement in this area.

1.3.2 Service Quality - Patient Expectations & Perceptions

Due to the intense competition in the healthcare industry, hospitals are focusing more on service quality as a means to maximize resource utilization and ensure their survival. Checking whether the service really meets the patient's expectations is what service quality is all about. In keeping with this, Gronroos said that patients judge the quality of service by comparing it to their expectations. In order for hospitals to enhance their services, it is crucial that they do quality assessments to discover what is standing in the way of quality development. A number of studies have shown a correlation between satisfied customers and high-quality service, which has prompted healthcare facilities to evaluate and enhance their service offerings. Thus, hospitals need to be inventive, quality-conscious, and use technically and highly competent staff as part of their real quality assurance initiatives to eradicate little or large deviations and achieve quality standards.

Furthermore, people often anticipate error-free treatment from hospitals, which is why hospitals frequently aim to provide services that fulfill established

quality standards. Nevertheless, it is crucial to consider patient expectations when evaluating the quality of these treatments, since they are not consistently met. A service is considered high-quality when it satisfies the customer's expectations and demands. What makes a service good is how it differs from what customers anticipate. The difference between expectations and perceptions is that the latter are based on the customer's prior experiences with the service, while the former are more directly tied to the present situation of service providing. Modern patients are able to assess a hospital's quality and compare its services to those of other similar facilities thanks to extensive online resources. Consequently, hospitals will gain an edge over their competitors by being well-informed on how patients see and anticipate their services.

1.4 Patient Satisfaction

The word "satisfaction" is used to describe when a need, want, or hunger is satisfied. How happy a patient is with the treatment they've had from their doctors and other medical staff is what we call "patient satisfaction" in the healthcare industry. It is a crucial indicator for evaluating healthcare facilities' level of care quality. When patients are happy with their treatment, it shows in more than just the results; it's a reflection of how they felt about everything from the kindness of the medical professionals to the cleanliness of the facilities.

Surgeon Donald Parsons, MD, of Denver, Colorado, eloquently summed up the many expectations related to patient happiness. People anticipate a complete recovery free of danger, difficulties, discomfort, and premature discharge from the hospital, he said. On top of that, you're required to provide outstanding care to the client as they undergo a series of procedures that are completely painless and risk-free: diagnosis, testing, probing, medication, and surgery. The complex and, at times, irrational expectations that patients have of their healthcare encounters are emphasized by this remark. The weight of these expectations, together with the need to provide safe and effective medical treatment, falls heavily on healthcare practitioners.[9]

The importance of patient happiness as a measure of healthcare service quality is becoming more acknowledged. This metric is important since it may help pinpoint specific places where things can be improved. Better patient outcomes and

experiences are the end result of using patient discontent as a tool to improve service delivery. There are many advantages for the healthcare business when patients are satisfied, according to several researches. Some of these advantages include a more loyal customer base, happier employees, more stable profits, less vulnerability to price competition, and better patient retention rates.

1.4.1 Importance of Patient Satisfaction in Healthcare

When it comes to healthcare organizations' reputation and performance, patient happiness is king. According to Oliver (1997), the feelings and anticipations around an event, such as going shopping, lead to a general psychological state known as satisfaction. This idea encompasses the whole patient journey in healthcare, from expectations before the appointment to the quality of treatment received during the visit and any subsequent follow-up encounters. Kim et al. (2008) further on the idea of patient satisfaction by stating that it is the evaluation of perceived value and the persistent reaction to service-related stimuli before, during, or after a patient consumes medical services. According to this notion, it's critical to know what patients want and how those wants affect their level of pleasure. Meeting clinical requirements is important, but ensuring patient happiness also requires attending to the sensory and emotional dimensions of treatment. As a measure of service quality, patient happiness is crucial, and healthcare practitioners must recognize this. Positive patient feedback is priceless to healthcare providers. Patients who have positive experiences are more likely to follow their treatment programs, develop lasting connections with their healthcare professionals, and promote the institution to others. All of these things help healthcare organizations thrive and endure.

1.4.2 Benefits of High Patient Satisfaction

There are many advantages, both for the company as a whole and for individual patients, to healthcare providers that put an emphasis on patient happiness. Among these advantages are:

1. **Customer Loyalty:** People are more inclined to stay loyal to a healthcare practitioner or institution if they are satisfied with their treatment. In the highly competitive healthcare sector, this loyalty is invaluable since it leads to recurrent visits and long-term connections.

2. **Improved Patient Retention:** Patient retention rates are higher when they are very satisfied. When patients are treated with respect and kindness, they are less inclined to go elsewhere for treatment, regardless of how good the alternatives are.
3. **Reduced Vulnerability to Price Wars:** People are less inclined to make healthcare choices based on price alone if they are happy with the treatment they get. As a result, healthcare organizations are less likely to be affected by price-based competition.
4. **Consistent Profitability:** Consistent income is a result of happy and loyal patients. Profitability may be further increased by the acquisition of new patients through favorable word of mouth recommendations given by existing patients.
5. **Increased Staff Morale:** Creating a pleasant work atmosphere is centered on ensuring patient happiness. Staff morale and willingness to provide high-quality treatment are boosted when patients express gratitude and offer good comments.[10]

1.4.3 Strategies to Enhance Patient Satisfaction

Healthcare providers who want their patients to be happy with their treatment should tailor their approaches to meet each individual's unique requirements. Some of these approaches include:

1. **Patient-Centered Care:** All healthcare choices and activities must revolve around the patient if a patient-centered approach is to be adopted. Recognizing and catering to each patient's specific tastes, requirements, and principles is at the heart of this method.
2. **Effective Communication Training:** Interactions between healthcare providers and their patients may be greatly enhanced by providing them with communication skills training. Improvements in the provider-patient connection may be achieved via the use of strategies including cultural awareness, active listening, and empathy.
3. **Feedback Mechanisms:** Healthcare businesses may get valuable insight into patient satisfaction by instituting methods for gathering input, such suggestion

boxes and questionnaires. Taking this criticism into consideration shows that you are dedicated to become better.

4. **Streamlining Administrative Processes:** Patients' experiences may be improved and their frustrations reduced by streamlining administrative processes including appointment scheduling, invoicing, and insurance processing.
5. **Leveraging Technology:** The use of patient portals, telemedicine platforms, and electronic health records (EHRs) all work together to make healthcare more accessible and convenient for everyone.
6. **Staff Recognition Programs:** A culture of quality and accountability is reinforced when healthcare professionals are recognized and rewarded for their contributions to patient satisfaction.

1.4.4 Factors Influencing Patient Satisfaction

Many things affect how patients feel about their treatment and how satisfied they are with it as a whole. The following are critical considerations: timeliness, appropriateness, respect, concern, patient safety, continuity, efficacy, and availability. The following is a detailed analysis of these factors:

1. **Appropriateness:** When it comes to healthcare, appropriateness is all about how well the treatments are tailored to each individual patient. Correct diagnoses, individualized treatment programs, and condition-specific care are all part of this. Appropriate treatment increases the likelihood that patients will feel heard and understood about their health issues. When patients feel that the treatment, they are receiving is either irrelevant or insufficient, it may lead to dissatisfaction.
2. **Timeliness:** Delays in treatment may cause aggravation and negative health effects, making timeliness an essential component of patient satisfaction. Appointment scheduling, test result retrieval, and treatment are all areas where patients appreciate speedy service. Patients may be less satisfied if they have to wait a long period for elective operations, outpatient clinic visits, or emergency department visits. The patient experience is improved when procedures are simplified and treatment is delivered efficiently.

3. **Respect and Concern:** Patient satisfaction is greatly affected by the level of respect and concern shown by healthcare personnel. Kindness, understanding, and respect shown to patients is much appreciated. Building trust and relationship with patients is the goal of healthcare providers who listen attentively, resolve problems, and include patients in decision-making processes. As a result of the empathy and compassion shown to them, patients report higher levels of satisfaction with their treatment.
4. **Safety of the Patient:** Ensuring patient safety is of utmost importance in healthcare and has a direct bearing on patient happiness. Patients anticipate a risk-free setting where they may undergo treatment without fear of infection or other complications. It is important to prioritize the well-being of patients by implementing comprehensive safety standards. These protocols include correct drug delivery, good cleanliness practices, and adherence to surgical checklists.
5. **Continuity:** When patients get their treatment from many physicians in a coordinated manner, this is called continuity of care. To make sure their medical history and treatment plans are understood and handled properly, patients appreciate regular contact and follow-up. Patient discontent and shattered perceptions of care may result from continuity disruptions such frequent clinician turnover or ineffective communication across departments.
6. **Effectiveness:** A healthcare intervention is considered effective if and only if it produces the intended health results. If a patient's symptoms improve, their recovery period shortens, and their general health improves, then the treatment was beneficial. Higher levels of patient satisfaction are associated with healthcare practitioners that administer treatments based on evidence and efficiently track progress.
7. **Efficacy:** The capacity of an intervention to bring about the desired outcomes under ideal circumstances is the primary emphasis of efficacy, which is closely connected to effectiveness. Treatments and procedures that have a track record of success and provide quantifiable advantages are highly valued by patients. Improved patient satisfaction may be achieved when healthcare practitioners

inform patients about the anticipated results and the scientific foundation of the therapies that will be prescribed.

8. **Availability:** The accessibility of healthcare services is a key factor in determining the level of pleasure felt by patients. From making an appointment to talking to a doctor to getting their hands on prescriptions, patients want their medical care to be accessible and straightforward. The accessibility of services is affected by factors such as personnel levels, geographic location, and resource allocation.

1.4.5 Principles of Patient Satisfaction

Here are a few concepts that help ensure patient satisfaction:-

Quality of Care

Healthcare organizations establish trust and loyalty from their patients based on the quality of service they provide. Accurate diagnoses, successful treatments, and fast and safe execution of all medical operations are the hallmarks of high-quality care. In addition to the medical experts' competence, patients evaluate the quality of treatment based on how empathetically they were treated.

Healthcare practitioners who engage in evidence-based practice base their recommendations for patient care on the most recent scientific findings and established clinical standards. Healthcare cost and accessibility are also important factors. Customers appreciate affordable services that don't skimp on quality, and they anticipate prompt service without long wait times.

Integrating state-of-the-art medical technology and maintaining clean settings further elevate the impression of excellence. For example, patients have more faith when facilities are clean, equipment is sanitized correctly, and safety rules are followed. Staff members that engage in continuous professional development are better equipped to maintain the healthcare facility's commitment to offering patients top-notch treatment.

Managing Anxiety, Fear and Pain

Anxiety, dread, and discomfort may vary in intensity when patients visit healthcare facilities, and these factors can greatly impact their whole experience. A key element of patient satisfaction is managing these emotions, since they have a direct influence on how patients perceive their treatment.

Uncertainty over diagnosis, medical treatments, or possible results is a common source of anxiety. Providers of medical treatment may lessen the impact of this by building rapport with their patients and explaining procedures in simple terms. When dealing with anxious patients, it is essential for doctors and nurses to show compassion and reassure them. As an example, one way to alleviate stress is to give patients a voice in decision-making and let them know how their treatment plans are going.

Equally important is pain treatment. Medications, non-invasive procedures, and emotional support are all ways that patients anticipate their healthcare providers to help them feel better physically. Care and attention are given via individualised pain treatment programs. To further enhance patient results, it is recommended to include professionals like counselors or pain management consultants.

Better Communication

The cornerstone of any contact in healthcare is communication. Making ensuring patients feel heard and appreciated is all about communicating effectively, which in turn promotes understanding and trust. Some of the non-verbal components include listening attentively, making and keeping eye contact, and speaking in an empathic tone.

In order for patients to completely grasp their diseases and treatment choices, healthcare practitioners must explain complicated medical concepts in simple words. That means being forthright about the choices, risks, and advantages. Better medication adherence and patient satisfaction are outcomes of well-communicated instructions on medication, follow-up visits, and lifestyle changes.

Mechanisms for providing feedback also improve dialogue. Healthcare facilities may find ways to improve by giving patients a chance to voice their experiences and concerns. Patients' time and concerns are valued when they are kept informed about wait times, procedure changes, or billing processes on a frequent basis.

Personal Attention

To provide each patient the individualized treatment they require, healthcare providers must adapt their practices and procedures. Personal preferences, cultural sensitivities, and unique medical needs cannot be adequately met by a cookie-cutter approach. Healthcare practitioners demonstrate their values by customizing treatment plans to meet the specific requirements of each individual patient and are committed to providing exceptional care.

Using patients' names, keeping track of their preferences, and listening carefully to their concerns are all parts of this philosophy. Patients have a favorable view of healthcare practitioners who are patient, empathetic, and genuinely concerned about their health.

With the use of technology, one-on-one attention may be improved. One example is the use of electronic health records (EHRs), which enable doctor's access to detailed patient details and thereby provide better, more tailored treatment. Comfortable waiting rooms and private consulting rooms are only two examples of how patient-centered design improves healthcare facilities.

Staff Accountability

One important factor in patient satisfaction is the level of accountability among healthcare professionals. All healthcare practitioners must consistently demonstrate professionalism, ethics, and responsibility to earn the confidence of their patients. When people are held accountable, medical errors are less likely to happen and, when they do, they are dealt with quickly and effectively.

Accountability is enhanced by implementing open and transparent mechanisms for monitoring and assessing employee performance. Healthcare professionals maintain the highest standards of practice via continuous training and

professional growth. Better service delivery is the result when institutions promote a culture of responsibility and employees feel they can own up to their responsibilities.

When employees respond quickly to patient concerns, keep their word, and provide consistent service, patients feel valued. Because billing mistakes or scheduling difficulties may have a detrimental effect on patient satisfaction, administrative personnel is also accountable for these concerns.

Healthcare providers may promote accountability at all levels by establishing strong processes and keeping lines of communication open.

Sound Organizational Culture

The foundation of patient pleasure is a solid company culture. It includes the beliefs, norms, and practices that characterize how a healthcare facility provides treatment. Above all else, a patient-centric culture emphasizes cooperation, honesty, and continual progress with the well-being of patients as its top priority. The leadership of an organization has a significant impact on its culture. Employees are inspired to provide outstanding care when leaders put their well-being first, support their professional growth, and promote open communication. Further reinforcement of desirable behaviors may be achieved by acknowledging and recognizing exceptional achievement. To guarantee that all patients, regardless of their socioeconomic situation or family history, get impartial, high-quality treatment, ethical procedures are also a priority in institutions with strong organizational cultures. Ethical use of medical technology, clear billing, and compliance with rules are their top priorities. Patients are more satisfied and devoted to these cultures because they promote trust and credibility.

1.4.6 Ten Pillars of Patient Satisfaction

In her 1984 landmark book "The Ten Pillars of Patient Satisfaction," Wendy Leebov laid forth the idea of patient satisfaction and how it relates to healthcare quality in general. In order to build a healthcare system that is patient-centered and guarantees happiness, loyalty, and good results, each pillar is essential.

Leadership Vision and Commitment

The leadership of healthcare organizations is crucial in determining their goals and future courses of action. Strong leaders prioritize the needs of their patients and work tirelessly to ensure they get high-quality treatment and are completely satisfied with their experience. To improve patients' experiences, leadership must establish priorities and objectives that are in line with the institution's purpose. A culture of responsibility and greatness is also encouraged by strong leadership. There is a domino effect that affects the whole healthcare company when leaders are involved with their employees, promote open communication, and set a good example. Leaders that put patients first spend money on education, tools, and facilities that boost service and care quality.

Process Design and Continuous Improvement

The smooth provision of healthcare services depends on efficient procedures. In order to increase productivity, process designers streamline processes, eliminate unnecessary steps, and apply best practices supported by evidence. On the flip side, continuous improvement refers to the persistent endeavor to pinpoint shortcomings and enhance these procedures for the benefit of patients. Process design is also heavily dependent on technology. Patients may benefit from better coordination, fewer mistakes, and a more convenient and individualized experience via the combination of automated systems, telemedicine, and electronic health records (EHRs).

Employee Engagement and Empowerment

Employees that are engaged and empowered play a crucial role in ensuring patient satisfaction. It is more probable that staff workers will provide outstanding care when they are driven and respected. The key to an engaged workforce is a positive work atmosphere that encourages employees to speak out, own up to their responsibilities, and work together productively. Stepping things up a notch, empowerment gives workers the freedom, knowledge, and resources they need to make choices that are good for patients. Better interactions with patients are another benefit of a healthy work environment that prioritizes employees' happiness and

work-life balance. When workers are happy in their jobs, they're more inclined to go the extra mile to make patients feel loved and valued.

Accountability for All Categories of Staff

Everyone on staff, regardless of their position, should take pride in their work and do their part to make patients happy. This include both clinical and non-clinical workers, including physicians, nurses, administrators, housekeepers, and security guards. Establishing clear goals, performance measures, and feedback systems is crucial for healthcare companies to create accountability. Individual or group weaknesses may be better understood with the use of frequent assessments, peer reviews, and patient comments.

Accountability also includes providing care with empathy and honesty, keeping patient information private, and following all applicable ethical guidelines. Trust in the healthcare system and satisfaction with treatment are both enhanced when all personnel adhere to these standards.

Monitoring and Feedback

In order to achieve continual improvement, it is necessary to monitor and provide feedback. Surveys, focus groups, and other real-time feedback tools should be routinely used by healthcare institutions to gauge patient satisfaction. By using these instruments, we can learn more about the things that patients care about and where we can make changes. Quantitative and qualitative information should be included into feedback systems. For instance, patients may express personal experiences via open-ended questions, and surveys can gauge satisfaction with different parts of treatment. Furthermore, staff members are more likely to work together to resolve areas of concern when feedback findings are shared with them. Patients are more likely to be loyal and trusting of a business that values their opinions and takes their suggestions into consideration.

Vertical and Horizontal Communication

To make sure that everyone is pulling in the same direction, good communication is key inside healthcare companies. In contrast to horizontal

communication, which is cooperation between colleagues across divisions, vertical communication is the conveyance of information from management to employees. Employees are better able to grasp the organization's priorities, objectives, and vision when there is open and honest communication across levels of management. A feeling of belonging and commitment to the organization's goals may be fostered when leaders keep their people informed and include them in important decisions.

However, cooperation and coordination rely heavily on horizontal communication. For instance, better patient care, fewer mistakes, and less room for miscommunication might result from open lines of communication amongst physicians, nurses, and administrative personnel. Patients' needs are met in a more comprehensive manner when different departments work together, which improves results and increases satisfaction.

Staff Development and Training

Continuous learning and improvement are cornerstones of providing first-rate care and services. Employees who participate in continuing education programs are better able to keep up with the ever-changing landscape of healthcare. Technical and soft skills should both be included in training programs. Workshops on cutting-edge diagnostic methods, for example, might be useful for healthcare personnel, while classes on customer service could help them engage with patients better. The ability to manage various patient groups successfully is enhanced when staff members are trained in emotional intelligence, cultural competency, and conflict resolution. Another factor that contributes to worker happiness and retention is the availability of mentorship programs and career development possibilities. Employees' dedication to their jobs and the pleasure of patients are both affected by the visibility of advancement prospects inside the company.

Reward and Recognition

Staff members that go out of their way to make patients happy are recognized via reward and recognition programs. An appreciative culture may flourish and people will strive to do their best when they are acknowledged for their contributions to care delivery, collaboration, and innovation.

Money, promotions, and special events to show thanks are just a few examples of the many shapes that rewards may take. Recognizing and rewarding progress toward goals, such as high patient satisfaction ratings, on a regular basis helps keep the focus on the organization's purpose and principles. Staff members are more likely to stay involved and dedicated to delivering outstanding service when they feel valued and appreciated.

Service Recovery

Healthcare businesses strive for excellence, yet delays, misunderstandings, or unfulfilled expectations may still lead to unhappiness. The goal of service recovery is to transform unfavorable experiences into good ones by quickly and efficiently responding to and resolving patient issues. A successful plan for restoring service requires admitting fault, offering genuine apologies, and acting swiftly to fix the problem.

In order to successfully restore services, it is crucial to train workers to address complaints professionally and with sensitivity. Stronger ties with patients and loyalty are fostered when organizations continually transform unhappiness into pleasure.

Patient Focus and Sustainability

One of the 10 pillars is patient attention, which states that patients' wants, needs, and health should take precedence above anything else. Listening carefully to patients, including them in decision-making, and customizing treatments to fulfill their unique needs are all parts of patient-centered care. Keeping patient satisfaction at a high level over the long term is what we mean when we talk about sustainability in this setting. Infrastructure, personnel development, and technology developments must be continuously invested in for this to be achieved. It also necessitates dealing with more general concerns, such environmental sustainability, which is becoming more important to patients who prioritize green medicine. Healthcare businesses may guarantee long-term happiness and loyalty by coordinating their operations with the values and expectations of their patients. Patients aren't the only ones who gain from sustainable practices; the company as a whole benefits from them as well.

1.4.7 Patient Satisfaction and Hospital Services

When it comes to healthcare facilities, patient happiness is measured by how well the services met or beyond their expectations. Patient satisfaction is a measure of how well healthcare met their needs, how well the system met their expectations, and how much they enjoyed their experience overall. It stands for harmony between how patients experience treatment and what they anticipate from it. There are many facets to health care, and this multi-faceted approach takes them all into account. Also, because to easy access to relevant information, today's patients are more informed and have higher expectations of healthcare systems. Consequently, fixing problems with service delivery is crucial now more than ever.

When patients have a positive outlook, they are more likely to have good results. Customers who have a positive experience at a hospital are more inclined to recommend it to their friends and family. However, poor adherence and, in extreme situations, unfavorable word of mouth may result from patients' negative attitudes and discontent with the hospital's health services, which in turn can deter others from obtaining treatment from the hospital.

Consequently, patient satisfaction is highly valued since it is associated with greater knowledge and retention of hospital information, more promptness in seeking treatment, and higher adherence to medical orders. Positive health results are guaranteed by all of these. Another important metric for evaluating a hospital's healthcare quality is the level of customer happiness. Health care delivery and patient-initiated delivery may both be enhanced via its evaluation.

1.4.8 Patient Satisfaction and Service Quality

Patients seek medical care with the sole intention of improving their health, which in turn will raise their productivity and, ultimately, their nation's economic development, regardless of the circumstances or conditions in which they get treatment. Consequently, healthcare plays a key role in national development by enhancing the well-being of the people. When it comes to connecting patients, doctors, and hospitals, health marketing is where it's at. It is widely believed that customer satisfaction is a major component of patient loyalty, good word of mouth, and repeat visits, all of which are important to the hospital. Patients' opinions,

especially on the quality of care, may influence their confidence and their actions when it comes to selecting a hospital. Medical treatment, administrative service, and nursing care are the three pillars upon which the excellence of public hospital services rests. There is no correlation between the infrastructure and interaction of public hospitals and the quality of care they provide. Doctors, nurses, and information all have a big role in how satisfied a patient is. In order to gain their loyalty, providers must meet their patients' expectations and demands. Regular communication between physicians and patients is essential for clinicians to grasp patients' needs and expectations and earn their loyalty. A patient's level of satisfaction with their healthcare treatment is defined as their subjective evaluation of that service. A patient's overall satisfaction with treatment was much enhanced by the physicians', nurses', and pharmacists' ability to explain things clearly and communicate well with one another. When evaluating the efficacy of a system for providing services, patient happiness is considered a key performance indicator. Opportunities for improvement, such as the strategic formulation of healthcare plans that go above and beyond both benchmarks and patient expectations, may be found in patient evaluations of treatment. The results of patient satisfaction surveys, however, are crucial for healthcare quality improvement initiatives.

Providers in the public health sector that want to see an uptick in service quality and happier patients should steer clear of making promises they can't keep and instead focus on creating standards for their services that encourage consistent reliability. In order to boost service quality and enhance consumers' impressions, public healthcare providers should create well-planned plans. Private hospital patients report higher levels of satisfaction and loyalty compared to their public sector counterparts.

Retention rates, therapeutic results, and medical malpractice lawsuits are all impacted by patients' levels of satisfaction. A key performance factor for healthcare providers and institutions is the level of patient satisfaction. The result of patient satisfaction is most strongly influenced by the doctor-patient contact. Health outcomes and treatment adherence may be greatly improved if doctors work on their interpersonal skills. Satisfied patients are more likely to stick with their treatments.

To realize its overarching policy goal of universal health coverage by the year 2000, the World Health Organization established 38 objectives in 1984. Goal 31 states that all member States shall have established reliable systems to verify healthcare quality for patients by the year 1990.

One of the many aspects of a healthcare system's service quality is the level of satisfaction its patients feel their needs and expectations are being met by the treatments they get. Furthermore, while assessing the quality of services and treatment provided, patient satisfaction is a crucial performance metric. The success or stagnation of a healthcare facility is directly proportional to the quality of the services it provides to its patients, making service satisfaction and quality an essential component of healthcare management.

When patients are satisfied with the care they get, they are more likely to remain loyal to the practice. The healthcare facility stands out in the competitive healthcare market because to its excellent service quality, which enhances patient happiness. Finding areas of improvement, outlining metrics for service measurement, tracking treatment and process results, and measuring patient satisfaction are all ways to raise the bar on service quality. In addition, the difference between intended and perceived service quality may be measured using variation analysis.

Appropriate assessment of healthcare services may better execute the delivery of high-quality healthcare services considering patients' perceptions. It has been noted that care providers often encounter challenges while trying to comprehend patients' decision-making processes. Reason being, personal experiences have a role in determining which hospital and its services a patient ultimately chooses. The patient's mental state, life events, and past experiences are the only factors that may influence how they perceive healthcare services.

Quality of service and patient happiness are the two most essential aspects of healthcare that influence a patient's decision on which hospital to choose. Every hospital has its own system in place for measuring service quality; nevertheless, the same SQ dimensions may be used to operate on patient happiness as well. Parasuraman et al. (1985) found that when patients think that the treatments they received were of good quality, their degree of satisfaction increases.

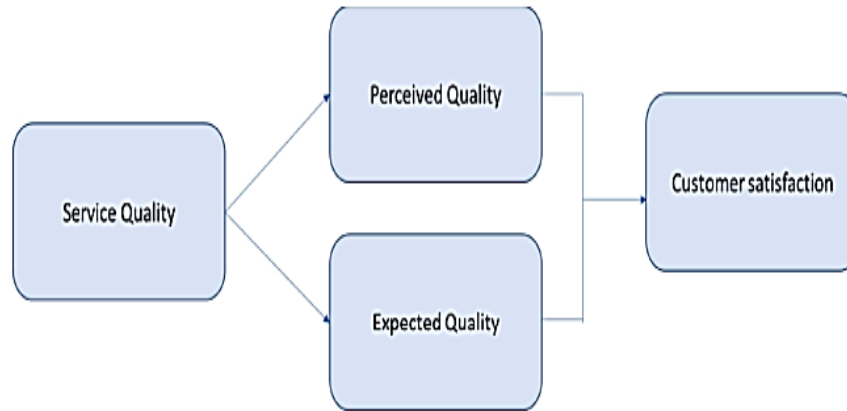


Figure 1.1: Service quality – expectations and perceptions

Sources: Secondary Sources taken from:

https://www.researchgate.net/figure/Estimation-of-structural-equation-model-parameters_fig2_363151788

1.5 Big Data Analytics

Analyzing massive and complicated datasets, or Big Data, for previously unseen relationships, trends, patterns, patterns, and insights is the goal of Big Data Analytics. "Big data" refers to massive datasets tested using a wide range of statistical methods and computer models. [11]

1.5.1 Overview of Big Data

There are a variety of approaches to processing structured, semi-structured, and unstructured data. A number of academics stress the importance of unstructured data in predictive analytics. Hadoop, Spark, and Map-Reduce are some of the methods used to address the problems with unstructured data in big data analytics, which Khaleel & Al-Raweshidy (2018) investigate. Adopting the six V's as outlined in the literature, it also tackles the problems with big data. What follows is an explanation of the power and significance of unstructured data in relation to predictive analytics. The research found that the model will now take unstructured data into account. Structured data follows a predetermined pattern. From the vantage point of big data analytics, both structured and unstructured information may be efficiently handled. As a result of these challenges and the clear links seen in large, high-dimensional datasets, deep

learning and machine learning have emerged as viable alternatives. Apache Spark employs a distributed ML library called Spark MLlib to facilitate the resolution of ML issues that need repeated processing. By combining the strengths of deep learning with machine learning, we can speed up computation for any kind of data. To deal with datasets with a lot of dimensions, machine learning techniques are used in conjunction with Spark-based models.

1.5.2 Nature of Analytics in Big Data

Big data analytics include a wide range of techniques, including classifications and predictive analytics.

Predictive analytics

When it comes to big data analytics, predictive analytics is only one kind. Using a wide range of methods, including deep learning, machine learning, and other artificial intelligence models, predictive analytics examines preexisting data. Predictive analytics provides strategies for dealing with high-dimensional datasets and both organized and unstructured data, two forms of big data. A deep learning framework for predictive analytics and structured data LSTM networks are described by Chen et al. (2016). Given that lung disorders were the primary emphasis, many of the methods were examined in light of predictive analytics as it pertains to lung diseases. Based on a number of parameters, including lifestyle, family history, smoking habits, and health issues including high blood pressure and diabetes, Muhammad et al. (2018) used structured and unstructured data to effectively predict outcomes. We cover a lot of additional ground when it comes to Hadoop-based predictive analytics. Some programmers are able to create Hadoop analytics. One such powerful predictive analytics tool is IBM SPSS Modeler, which offers predictive algorithms and an intuitive graphical user interface. It takes bigger datasets as input, converts them into streams, and then transmits the information in Analytic Server Language (ASL) to the IBM Analytic Server, which is running on the Hadoop environment. The design chooses to communicate analysis via spark or the local system if the storage needs are big enough. In the event that spark is not accessible, the map-reduce paradigm is used. Through the use of various data mining techniques, it is possible to forecast future health issues. A tool that processes huge data sets simultaneously while using mining

methods would be more effective than others, despite the fact that data mining improves accuracy and state-of-the-art illustrates performance.

It is preferable because using parallel processing or prediction speeds up management processes and saves time. A model has been put into action. Using a Hadoop-based map-reduce approach, they add a traffic prediction model called TFPC. They built the model using two modules, ODT and OPP, which stand for offline distributed training and online parallel prediction, respectively. Furthermore, the classification method makes use of correlation data derived from a parallel closest neighbor optimization classifier, which pertains to traffic flows. Lastly, ODP measures the outcomes and current data in conjunction with ODT's results to improve traffic flow prediction. Other performance indicators are proven by this unique methodology that blends map-reduce with mining methods. Among the top applications built using Spark in big data, one that would be useful for consumers of big data would be a real-time application that predicts how long patients would have to wait for treatment.[12]

The suggested PTTP algorithm estimates how long a patient will have to wait for each treatment task. They examine the waiting time for each job by collecting actual data from different hospitals. Depending on the likelihood of waiting, a Hospital Queuing Recommendation (HQR) determines the necessary course of therapy for patients. Even more so, in a cloud-based Apache Spark setting, the PTTP algorithm and the HQR system plot a strategy for an effective and quick response method. Spark applications are widely used, making it easy and simple to forecast any queuing systems.

Classification

In classification, the focus is on techniques and algorithms that can organize data that is not yet organized into predefined categories. In addition, it offers appropriate solutions for improved decision-making. Classification consists of two stages. The first involves learning strategies that equip the training set to learn with several prospects. The assessment of the training and test sets is the second stage. But now, parallel classification is the way to go for most large data classification issues. Two examples of classification techniques are machine learning and deep learning.

1.5.3 Classification of Big Data

The various formats for Big Data categorization are shown in Figure 1.2.

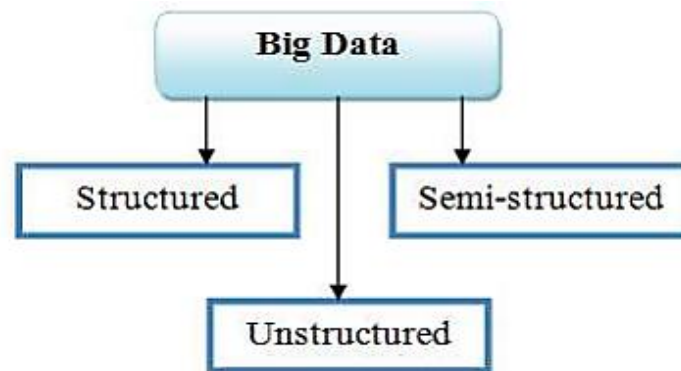


Figure 1.2 Classification of big data

Sources: Secondary Sources taken from:

<https://www.realcode4you.com/post/what-is-abstract-data-type-adt-in-java-data-structure>

Structured Data

The database already contains structured data, which is intended for its fixed format. All of the learning and computer-related pursuits make use of these formats. It is possible to explain structured data using human beings, expert systems, or other machine-centric sources. Data pertaining to machines may also be found in medical records, server logs, and other massive datasets from many sources. Examples of human-generated data include biodata, biographies, and linkages constructed from other sources that include personally identifiable information.

Unstructured Data

Unstructured data lacks any kind of organization, as contrast to structured data that is stored in a database or has a predetermined format. Of all the data that exists, only 20% is in an organized format. Unstructured data refers to the remaining 80% of data. There are two methods for processing unstructured data: expert-generated and system-generated. Information recovered from satellite photos, weather prediction images, radar images, and other similar sources is comparable to system produced

unstructured data.

Unstructured data created by experts has been accessible globally. Social media sites like Facebook are examples of user-generated, unstructured data and Twitter as well as information that is submitted via human interaction. Media files such as photos, videos, audio files, and satellite pictures are examples of unstructured data.

Semi-Structured Data

Data is presented in two formats: structured and unstructured. The former is arranged with certain qualities for processing, while the latter is semi-structured. Not all data is organized; some of it lacks the formality of relational databases or other data tables, but it does have certain characteristics that make it easier to analyze, such as semantic tags.

Since HTML does not limit the amount of data that may be included on a page while yet preserving categorization via syntactic information, it is an excellent example of semi-structured data. What we call "big data" is really structured information that defies easy categorization. Even while semi-structured data doesn't fit well into relational databases, it does have a number of desirable qualities that make it an excellent choice for analysis. Procedures having several stages may be stored in a relational database.

1.6 Machine Learning

The high dimensional data seen in areas such as cardiology, pathology, and radiology makes them promising topics overall. Trained machine learning can analyze photos, detect abnormalities, and identify problematic areas, improving the accuracy of all these operations. Machine learning has the potential to enhance healthcare performance by offering objective evaluations and insightful analyses. Using a unique technology, Health Catalyst analyzes data and returns it to clinicians in real-time to aid in clinical decision-making. In addition to storing vital medical records, electronic health records (EHRs) help doctors diagnose patients, prescribe tests, and suggest preventative screenings based on symptoms and other data.

1.6.1 Machine Learning Methods

There are primarily three ways to train ML models: supervised learning, unsupervised learning, and semi-supervised learning.

Supervised learning

The use of human-provided labels in supervised learning implies the use of training layers. In order for learning algorithms to make predictions about the output, they train the layers using labels and training mistakes. Furthermore, it anticipates labels by analyzing patterns. In supervised learning, for instance, the input vehicle is used to train and assess labels in order to eliminate mistakes. Then, the model is trained to predict the automobile. One common use of supervised learning is the prediction of stock market movements via descriptive statistics.

Unsupervised learning

To train an algorithm in unsupervised learning, one must not be aware of the labels or the data format. Using the input from earlier layers, it learns and makes predictions about the output. In unsupervised learning, the goal is to find all hidden patterns and use them to train the layers, so that the system can autonomously calculate using the representations it has found.

Transactional data often makes use of unsupervised learning. items with a certain scent are likely to target customers aged 20–40, but with an endless database and no idea what kinds of items or comparable features may be retrieved from consumer data, it's hard to tell. When it comes to detecting anomalies, such fraudulent card transactions, unsupervised learning has shown to be a powerful tool. Additionally, it foretells potential crimes on certain websites, recommends things to purchase next, and much more besides.

Semi-supervised learning

A few of the most popular types of learning include supervised learning, which uses human-provided labels to train the input layer, and unsupervised learning, which uses feedback from earlier layers to teach subsequent ones. Lacking labels and a framework for training and generating algorithms, it differs from supervised learning.

Combining supervised and unsupervised learning techniques, a semi-supervised learning system may acquire new knowledge. What this means is that there is both labeled and unlabeled data present. Just think about how the output might be predicted if labels were applied to images: they would follow certain established patterns. Subtle patterns and feedback from earlier layers may be predicted using unlabeled photos.

1.6.2 Machine Learning Data Types

Raw data would be acquired at the input layer. Structured and unstructured data both make up the data at the recognized tiers. Numerical data and categorical data are the two main components of structured data.

Continuous data

These numbers fall into the category of numerical data, which can be analyzed using tools like standard deviation and mean with intervals. The results of statistical tests like the t-test, Pearson correlation coefficient, and F-test might provide light on the data. Factors such as a person's weight and height, loan interest rates, etc.

Discrete data

Discrete data is data that has integer values. Additionally, it uses probability to determine when occurrences take place. Data that is floating or decimal is mostly out of its league. Think of a class of students or a rainbow of colors as examples.

Nominal data

Data that has predetermined labels or structures is called nominal data. For statistical purposes, it is unnecessary to calculate things like mean, mode, and standard deviation. For instance, a country's states or a region's zip code.

Ordinal data

The information is organized into categories. Ordinal data analysis is shown by calculations using a non-parametric test. In addition, the standard deviation, mode, and mean of the data do not give any useful information.

Binary data

Examples of recognized binary data include gender (male or female), fraudulent transaction (yes or no), and cancerous cell (true or false). This form of data is a specific instance of nominal data, which takes the values true (1) or false (0).

1.6.3 Machine Learning Algorithms for Healthcare Data Analysis

"Machine Learning" in 1959 had the ability to play chess and other strategic games. Machine learning may be defined as a system that enables computers to autonomously acquire new knowledge, without human intervention. Creating an application that can access and utilize data for learning is the primary goal of machine learning. Primarily, it relies on probability and statistics.

Nevertheless, when it comes to computational decision-making, it outperforms more conventional statistical methods. The foundation of effective artificial intelligence is machine learning. Many fields can benefit from it, including medicine, engineering, and society at large.

The foundation of machine learning is data. When healthcare providers have access to relevant data, clinical judgments may be improved. You can't have a model without it. Machine learning algorithms are used to derive valuable insights from healthcare data, enabling the making of superior clinical diagnostic judgments. Until now, several well-known learning methods have been developed, including Naïve Bayes Support Vector Machine, K-Nearest Neighbor, Neural Network, Decision Trees, and many more.

K-Nearest Neighbour (KNN)

One of the most basic supervised learning algorithms, the KNN handles both regression and classification, two types of predictive problems. K-Nearest Neighbors (KNN) is one of the simplest machine learning algorithms. The key concept of the KNN approach is simple. In fact, the class of an instance's closest neighbors determines their order. The technique is known as KNN because in this case it might be useful to take into account more than one immediate neighbor (let's call them K). The computational cost of running KNN is high since it requires a lot of memory for

the testing and training sets.

To be more precise, we find the separation between the visible and unseen instances by using the distance measure approach. As far as distance measurements go, the one that stands out most is the one based on the Euclidean formula. By picking the visible examples that are farthest away from the invisible one, we may use the distance metric to categorize the unseen cases. There are basically two stages to K-Nearest Neighbor. We start by picking the most pleasant people to live next to, and then we use the distances between them to determine which class to put them in. Notably, whereas the mean value is used for regression, the mode value is utilized for classification.

Support Vector Machine (SVM)

Classification and prediction are two areas where support vector machines (SVMs) shine. Both linear and non-linear classification problems are good candidates for this approach. Classification of non-linear data is often accomplished using this method. In order to transform the initial training data from a low-dimensional space to a high-dimensional one, support vector machines use a non-linear function called a kernel function. To further divide the training set into two categories, it finds the linear optimum hyperplane to use as a decision boundary.

Similar to a tuning parameter, the kernel function is used in support vector machine models. With the help of the kernel function, the separable issue is transformed from non-linear to separable. When using the kernel function to divide data into two categories, a hyperplane is always an option. The SVM finds this hyperplane for data separation by employing support vectors, which are a small subset of the training data (feature vector).

Linear, binary, and non-probabilistic instances of unseen data are classified using the SVM classifiers. An efficient binary classifier for dealing with the multiclass problem, pair-wise classifications are time-consuming yet effective. Support vector machines (SVMs) are often called "black boxes" because of the high number of unknown properties that make them challenging to examine, comprehend, and predict.

Decision Tree (DT)

When representing a classification model, DT is a crucial machine learning method. A decision tree (DT) is a model of a tree with nodes, branches, and leaves. A feature (or attribute) is represented by each internal node in this model, a decision (or condition) by each branch, and a class (or result) by each leaf node. The tree's root node figures out how to recursively split it up according to the values of its attributes. There are two stages to Decision Tree (DT): construction and trimming. Attribute selection measures (ASMs) including the Gini index, the gain ratio, and the information gain are used to choose features at each level during construction [9], and outliers are deleted during pruning. When dealing with tiny datasets for classification, DT approach works well. The technique is easy to grasp, however it has issues including biased findings and overfitting when dealing with unbalanced datasets. When building decision trees (DTs), three popular methods are ID3, C4.5, and CART.

Naïve Bayes (NB)

NB is a way to learn classifications that uses Bayes' theorem. Mathematical statistics and probability theory form the basis of the Bayes theorem. For that reason, probabilistic learning classifier is the name given to Bayesian classifiers. Statistical classifiers are actually its formal definition. Binary and multiclass classification problems are both handled by NB.

Based on the provided class label, the Naïve Bayes (NB) classifier calculates the class membership probability. The premise of an independent predictor is Naïve Bayes' primary shortcoming. No two features (attributes) of the same class are dependent on each other in this approach. Although it has some similarities with Support Vector Machine (SVM), Naïve Bayes makes use of statistical methodologies.

Neural Networks (NN)

For Neural Networks (NN), the supervised learning procedure is the bedrock. One of its famous applications is fixing problems with classification. Neural networks are artificial structures that machine learning uses; they are designed to function similarly to the human nervous system. In 1943, McCulloch and Pitts created a neural network, a mathematical and algorithmic model for computers. "Neural networks" are

mathematical models that attempt to simulate brain function. Think of a directed network with three nodes—input, sink, and internal, sometimes called hidden—to see how this method works. One layer contains the nodes that feed data into this model, while another layer contains the nodes that send data out of the model. The remaining nodes are hidden in one or more levels. The edge displays the activation value. The three main components of a neural network are the input, hidden, and output layers. Every node in one layer is connected to every node in the other layers. Adding more hidden layers allows one to construct a deep neural network. The error-function, search-function, and update-function are the three main components of a neural network's normal functioning. Applying the error-function to a certain set of inputs allows one to ascertain whether the output is of high or poor quality. To reduce the number of mistakes, the search tool identifies the problematic locations. What needs to be modified will be decided by the update function based on the search results. Iteratively improving the algorithm's performance is the goal here.

Learning techniques based on Neural Networks (NNs) are very useful when dealing with complex problems that cannot be easily described using traditional software and logical analysis. Many models of neural networks are created for supervised classification. A large percentage of NNs use the backpropagation multilayer feedforward method.

Logistic Regression (LR)

Analyzing the link between two variables allows for their modeling and prediction using regression analysis. Both logistic and linear regression methods exist. While linear regression needs a continuous value, logistic regression (LR) needs a discrete one.

1.7 Problem Statement

The purpose of this research, which is titled "A STUDY ON ENHANCING PATIENT SATISFACTION AND SERVICE QUALITY THROUGH MACHINE LEARNING CLASSIFIERS AND BIG DATA ANALYTICS," is to investigate how advanced machine learning methods and big data analytics can benefit healthcare service quality and patient happiness. By examining massive healthcare datasets, such as patient comments, service delivery metrics, and clinical outcomes, the study aims

to discover critical components that impact patient experiences. The study's overarching goals are to improve healthcare operations, decrease patient wait times, and tailor treatment plans to each person's unique requirements via the use of predictive models and machine learning classifiers. Improving service quality and patient satisfaction levels is the end objective of proposing data-driven techniques that healthcare professionals may use to offer more efficient, patient-centered care.

1.8 Need and Scope of The Study

Need of the study

Research into healthcare delivery optimization in the current era using big data analytics and machine learning classifiers raises the bar for service quality and patient satisfaction. It is getting more and more difficult for healthcare providers to meet the diverse and ever-evolving requirements of patients without lowering the quality of service they provide. Patients' levels of satisfaction are indicators of both the effectiveness and efficiency of healthcare delivery systems in caring for individual patients. Conventional approaches to healthcare management could overlook certain dimensions of patient happiness. By combining machine learning classifiers with big data analytics, healthcare providers now have a game-changing tool at their disposal to sift through enormous datasets in search of previously unseen insights. Improving service quality, forecasting patient outcomes, and finding characteristics that impact patient satisfaction are all aided by this study. To add to that, as healthcare shifts to a more preventive and individualized paradigm, it is critical to comprehend and embrace cutting-edge technologies. Ultimately, the study aspires to improve healthcare delivery by enhancing patient satisfaction and service quality through practical insights that bridge the gap between current methods and the revolutionary possibilities offered by machine learning and big data analytics.

Scope of the study

This project aims to improve healthcare service quality and patient happiness by using machine learning classifiers and big data analytics. Patient demographics, service comments, medical records, and operational data are some of the massive datasets that this approach seeks to sift through in order to uncover trends and elements impacting patients' overall experiences. Predictive models for better resource allocation, shorter

wait times, and optimized service delivery procedures are part of the study. Personalized care techniques that make use of data analysis findings are also explored. Healthcare administrators, clinicians, and patients are all intended beneficiaries of the study's practical suggestions for enhancing service quality. In addition, the research emphasizes how new technology are changing healthcare operations, so the results may be used in a variety of healthcare settings (e.g., hospitals and clinics) and expanded upon for use in data-driven healthcare management in the future.

1.9 Objectives of The Study

Following are the main objectives of this study.

1. To discuss the importance of sentiment analysis in patient satisfaction.
2. To present a unique SCSP Ensemble Model for evaluating Patient Health Care Opinion Systems.
3. To provide a unique feature selection approach for determining the main feature in Home Health Care Services using patient satisfaction data.
4. To demonstrate a unique Sentiment Analysis technique for evaluating patient drug satisfaction by reporting symptoms
5. To compare the performance of the suggested methods to that of current strategies.

1.10 Definition of The Keywords

Patient Satisfaction

The term "patient satisfaction" describes how happy people are with the medical treatment they get. It includes a wide range of aspects, including the following: the ease of the healthcare facility, the time spent waiting, the quality of treatment received, and the communication between patients and doctors.

Service Quality

The term "service quality" is used to describe how well healthcare services meet or surpass patient expectations. The quality of care, how quickly it is provided, how easy

it is to get in touch with, and how satisfied patients are with their experiences all fall under this category.

Big Data Analytics

In big data analytics, huge and complicated datasets, or "big data," are analyzed for insights, patterns, and connections. To enhance service quality and patient happiness, this technology is used in healthcare to assess patient information, treatment effectiveness, and operational performance.

Machine Learning Classifiers

The domain of artificial intelligence makes use of algorithms known as machine learning classifiers to sort data into predetermined groups. Using training data, these classifiers improve healthcare decision-making, identify illnesses, and forecast patient outcomes.

Healthcare Operations

Managing and administering healthcare services, such as patient care, personnel, finances, and logistics, are referred to as healthcare operations. Service delivery, cost reduction, and patient happiness may all be greatly improved by healthcare operational optimization.

1.11 Limitations of The Study

- **Data Availability and Quality:** Complete, consistent, and impartial data is essential for machine learning models and big data analytics to be accurate. The dependability of the insights and projections might be affected by missing or low-quality data.
- **Privacy and Security Concerns:** Strict respect to privacy standards, such as GDPR and HIPAA, is necessary when dealing with patient data. It may be quite a challenge to ensure compliance while yet keeping data secure.
- **Complexity of Healthcare Systems:** There are a lot of aspects (care quality, hospital atmosphere, staff interactions, etc.) that affect patient happiness, and the models may not capture all of them. So, the research can't be considered complete.

- **Generalizability:** Results may not be generalizable to other healthcare systems due to differences in healthcare system resources, patient demographics, and operational methods.
- **Interpretability of Models:** Lack of transparency and difficulty in interpreting by healthcare practitioners without technical knowledge are potential barriers to practical deployment of machine learning models, particularly complicated ones.
- **Technological Constraints:** The study's potential limitations may be imposed by healthcare facilities that lack access to sophisticated computer systems, data storage facilities, and analytical tools.
- **Subjectivity of Patient Satisfaction:** Analytics struggle to capture all aspects of the patient experience due to the subjective nature of patient satisfaction and its impact by individual preferences and expectations.
- **Integration with Healthcare Systems:** It may be difficult to incorporate the results into current healthcare procedures, particularly in environments where there is a lack of technology uptake or aversion to change.
- **Resource Constraints:** Data collecting, model creation, and implementation may be time-consuming, expensive, and labor-intensive, which might restrict the study's scope.

1.12 Proposed Framework of The Study

Research Procedure

Using ideas from sentiment analysis and machine learning, this study proposes a new approach to hospital recommendation systems. Time, location, professional care, staff behavior, and financial feasibility are some of the factors that the end user considers while using this new hospital recommendation system. The thoughts of many end users from online hospital reviews provide the basis for the analysis of user comments.

In order to construct a realistic model and examine customer input before recommending hospitals to different healthcare businesses, the suggested approach

makes use of text mining techniques that take uncertainty into consideration. In order to investigate this matter, we provide a new ensemble method and feature selection algorithm that can effectively and reliably extract statements for healthcare recommendation purposes.

Data collection and pre-processing

Researching online discussion groups hosted by various healthcare institutions is the first stage. Python scripts are used to obtain the data from the Outscraper.com API. After entering the ratings into Excel, they are converted to CSV. We meticulously classify all reviews as positive, negative, or neutral, regardless of whether they were originally tweeted or not. We do not include reviews that are written by other parties. Feedback is gathered from six different hospitals that provide a range of medical specialties.

Proposed architecture

After removing unnecessary reviews, we include both new and old ones, which are then carefully rated as positive, negative, or neutral. The research takes into account the feedback provided by six hospitals.

Feature Extraction

For medication names and dosage contexts, the label encoder converts the categorised data into numerical data. Next, we run sentiment polarity analysis on the collected data. Both Independent Component Analysis (ICA) and Principal Component Analysis (PCA) are used in this section of the investigation. It is believed that the next study would bring about revolutionary changes in healthcare delivery by using machine learning classifiers and big data analytics to improve service quality and patient pleasure. The accuracy of determining factors impacting patient satisfaction may be greatly enhanced by using machine learning classifiers to historical data in order to spot patterns.

Consequently, these preventive actions are anticipated to substantially increase the overall level of patient satisfaction. Many people think that using predictive models made by machine learning algorithms would significantly enhance the quality of

healthcare services. Predicting patient outcomes, preferences, and possible repercussions will allow healthcare businesses to optimize resource allocation, expedite procedures, and improve treatment customization. Healthcare delivery will become more efficient and effective as a result of this.

Further, improved sentiment analysis of patient comments is expected to result from big data analytics, which could lead to a more thorough comprehension of patient experiences. Healthcare professionals may learn more about their patients' thoughts, feelings, and preferences by mining unstructured text data for valuable insights. As a result of this information, doctors will be able to personalize their care for each patient. The study's authors hope that, with the help of machine learning classifiers and big data analytics, patient-centered care will become the standard in healthcare in the future. Patient satisfaction, service quality, and proactive, individualised healthcare delivery will all see significant upticks with the shift to this paradigm.

1.13 Plan of Work

Chapter 1: Introduction

To set the stage for the research, this chapter explains why it's important to use technology to improve healthcare service quality and patient happiness. It draws attention to the increasing difficulty healthcare providers have in catering to patients' varied demands and the critical role that big data analytics and machine learning classifiers may play in finding solutions to these problems. This chapter also details the study's aims, scope, and methods, which were to investigate healthcare delivery, sentiment analysis, and the general effect of predictive modeling on patients' experiences with and satisfaction with their treatment.

Chapter 2: Review of Literature

The second chapter offers a thorough literature review on healthcare topics such as patient happiness, healthcare service quality, sentiment analysis, and the use of big data analytics and machine learning in healthcare. This article reviews the literature on healthcare delivery optimization, prediction models, and sentiment analysis for patients' perspectives. By reviewing several approaches, models, and results from pertinent sources, this chapter finds holes in the present literature and argues that this

study is necessary to fill those gaps.

Chapter 3: SCSP Ensemble Model and Feature Selection for Predicting Patient Healthcare Opinions and Home Healthcare Service

To forecast patients' healthcare-related views, this chapter presents the SCSP (Sentiment Classification with Semi-supervised Learning) ensemble model. Improving the accuracy and efficiency of models is another important topic covered in the chapter, along with feature selection strategies. Using information retrieved from reviews and comments, this model attempts to forecast how satisfied and what kind of opinion patients would have about home healthcare services. A successful prediction model for healthcare services is created by exploring key methodologies such as feature selection, machine learning classifiers, and text mining.

Chapter 4: Sentiment Analysis and Performance Comparison for Evaluating Patient Opinions on Drug Satisfaction

Using sentiment analysis to gauge patients' levels of satisfaction with their medications is the subject of Chapter 4. It evaluates the efficacy of several machine learning models for analyzing patient feedback on pharmaceutical items and looks at different sentiment analysis approaches. The purpose of this chapter is to find the best ways to analyze patient feedback in order to find out how satisfied people are with the drug's effectiveness, any adverse effects, and everything else that affects their experience.

Chapter 5: Conclusion, Recommendations, and Future Scope

In the last chapter, the investigation is brought to a close by reviewing the important points made and discoveries made throughout the research. The report delves into the results' policy and healthcare provider implications while also providing suggestions for enhancing service quality and patient satisfaction via machine learning and big data analytics. The chapter concludes by discussing the study's shortcomings and offering suggestions for further research, such as how to improve machine learning models, how to do sentiment analysis, and how to incorporate new technology into healthcare.

REVIEW OF LITERATURE

Alkhnabashi, Omer et al., (2024) [1] A essential, low-cost source of feedback for medical service management, online medical forums have evolved as places where people may discuss their experiences and seek advice. Not only does this information help measure patient satisfaction and improve healthcare quality, but it also sheds light on the effectiveness of pharmacological treatments, pain management strategies, and alternative therapies. This research meticulously catalogs important elements of patients' experiences, highlighting both the good and bad feelings conveyed in their accounts. Approximately 15,000 items were culled from different parts of the popular medical forum, patient.info, for our dataset. To examine this data, our novel method combines content analysis with deep learning approaches, aspect-based sentiment analysis, and LLM. Our approach is crafted to reveal many sorts of aspects expressed in patient feedback. Finding seven unique aspect kinds in the feedback, the investigation proved that deep learning models could accurately forecast the sentiment values associated with these aspect types. Notably, compared to other models, the LLM with few-shot learning fared better. Healthcare providers and medical service managers may benefit greatly from our research since it sheds light on patients' experiences in online forums and shows how useful sophisticated analytical tools are for gleaning insights from patients' unstructured input.

Okorie, Gold et al., (2024) [2] Big data analytics, a relatively new discipline, has opened up hitherto unimaginable possibilities for highly targeted advertising campaigns. By outlining the goals, methodology, important results, and conclusions from recent research in this subject, this review seeks to synthesise the existing knowledge on exploiting big data for customized marketing. Our major goal in writing this evaluation was to find out how marketers may use big data analytics to target specific customers based on their interests, habits, and other personal details. The research follows a thorough methodology, looking at several studies that use different big data tools and strategies for customized marketing. These tools and techniques include data mining, predictive analytics, and machine learning algorithms. Important results show that marketers' capacity to comprehend and

forecast customer behavior is much improved by big data analytics, which in turn leads to more efficient targeting and segmentation tactics. By sending more relevant and timely marketing messages, big data integration has increased consumer engagement, happiness, and loyalty. Data privacy issues, the need for sophisticated analytical abilities, and the possibility of inaccurate data are some of the obstacles that are brought to light. Research in the future should focus on creating frameworks for ethical data use in marketing practices and investigating how new technologies such as the IoT and AI affect customized marketing. This study provides useful insights for scholars and practitioners in the area of customized marketing by highlighting the revolutionary potential of big data.

Tanasescu, Laura et al., (2024) [3] Given the amount of changes that processes and people have encountered in many businesses in the last several years, it is essential for most firms to boost employee performance and productivity. Since this can maintain the company's development and competitiveness, it becomes even more crucial. Data pre-processing, variable selection, algorithm construction, and hyperparameter tuning are all part of the process of developing a machine learning algorithm to forecast performance scores for individuals in a single organization. Several approaches are discussed and contrasted in this study. Our primary objective in conducting this research was to identify the best methods for predicting the given variables, with the hope that subjective human judgment will have less of an impact on performance reviews and that overall productivity would rise.

Gupta, Arun. (2024). [4] As businesses want to boost growth and improve customer experiences, big data analytics in CRM systems are becoming more and more important. Several big data analytics techniques used in CRM are explored in this article. These include sentiment analysis, segmentation, text analytics, real-time analytics, social media analytics, predictive analytics, customer journey analysis, and social media analytics. Reviewing the pros and cons of integrating big data with CRM systems, this article follows the process. consumer satisfaction levels are higher, marketing campaigns are more targeted, and better consumer segmentation are all advantages. Topics covered include creating a data-driven culture in enterprises, addressing data privacy problems, and the need for data governance. The article shows how Big Data-driven CRM projects work in practice by referencing case

studies and examples from businesses such as Walmart, Amazon, Netflix, and Airbnb. These cases show how companies may streamline their operations, provide better customer service, and personalize user experiences with the use of big data analytics. Even with all the obstacles, Big Data analytics and CRM have a bright future ahead of them. They can drive innovation and change, which will deepen customer interactions and drive corporate success. The paper finishes by discussing potential trends in the future and providing suggestions on how CRM practices might successfully include Big Data analytics in order to compete in the data-driven economy of today.

Usman, Muhammad et al., (2024) [5] Since the COVID-19 pandemic began two years ago, public anxiety about health care systems has skyrocketed. In order to assess, track, and improve medical services, medical practitioners and allied organizations contact patients and ask for their opinion. Social media sites such as Facebook, Instagram, Twitter, etc. often include such opinions and perspectives. For researchers looking for up-to-the-minute news, commentary, and debate, Twitter is by far the most popular and widely utilized site. It is a great place to track public sentiment on different issues because of the viral content and popular hashtags (#). There are three hashtags used to extract the tweets: #healthcare, #healthcareservices, and #medicalfacility. Additionally, this research takes into account location and sentiment analysis in tweets. The accuracy of the outcomes is lower, according to a number of recent research that used Twitter datasets and ML and DL models. Furthermore, there was a lack of validation and thorough comparative analysis in the investigations. One of the research topics that this study seeks to answer is, "How do people around the world feel about their healthcare system?" and secondly, when it comes to healthcare twitter sentiment categorization, So far, how have deep learning and ML methods fared? An array of well-known ML models are used in the trials, including decision tree, AdaBoost, k-nearest neighbor, SVM, logistic regression, additional tree classifier, and Gaussian naive Bayes. Also, using the healthcare dataset, this research suggests an LSTM-ETC model that uses transfer learning to accurately forecast the customer satisfaction level. The suggested model surpasses the ETC model, which performed best with a precision score of 0.95, according to the results.

Amoo, Olukunle et al., (2024) [6] Across all industries, success in today's data-driven world hinges on companies' capacity to make good use of big data and analytics. This all-inclusive study delves into tactics for driving company growth with big data and analytics, highlighting important trends, obstacles, and best practices. Big data and analytics are introduced as a means for businesses to get useful information from massive data sets in the first part of the study. Data collecting, processing, analysis, and visualization are just a few of the big data and analytics tactics covered. The article covers important developments in analytics and big data, including the rising tide of AI and ML used to automate data analysis. Additionally, the paper discusses problems with big data and analytics and how to fix them, including issues with data security and privacy. Data quality, governance, and cross-departmental cooperation are highlighted as best practices for using analytics and big data to drive company success.

Raju, Nidadavolu et al., (2024) [7] With the ever-increasing complexity and amount of data, predictive analytics has become an indispensable tool for healthcare, banking, and e-commerce decision-makers seeking meaningful insights from big data. Different algorithms have different processing needs, scalability, and algorithmic performance; these factors make it difficult to choose the right machine learning algorithm for predictive analytics applications, particularly when dealing with massive data. This study presents a thorough evaluation of well-known machine learning algorithms used for predictive analytics, with an emphasis on how well they work and how feasible they are in large data settings. The research classifies algorithms according to their learning type (supervised, unsupervised, or reinforcement learning) and then assesses them on many criteria, including prediction accuracy, computing efficiency, scalability, and real-time analytics applicability. The merits and limitations of various algorithms, including as decision trees, linear regression, neural networks, support vector machines, and clustering approaches, are examined in depth to see how well they handle enormous datasets. To further evaluate the algorithms' predictive power and computational feasibility, the paper presents a set of criteria such training time, F1-score, and accuracy. This case study provides insights into the real-world performance of these algorithms in various predictive analytics situations by applying them to a sample large data collection. To better understand the costs and benefits of different algorithm options, Recommendations

for choosing machine learning algorithms according to particular predictive analytics goals, data properties, and processing needs are provided in the latter section of this work. It goes on to talk about how these algorithms can be hard to implement in big data settings, but it also looks at ways forward, like using distributed computing and integrating deep learning, that could improve predictive analytics in the future.

Nwaimo, Chioma et al., (2024) [8] The introduction discusses the importance of data analytics in healthcare and lays out the study's purpose. It highlights the evolution of data analytics in healthcare, the many types of healthcare data, and the challenges associated with collecting and maintaining this data. The next section will examine the role of predictive modeling in healthcare, focusing on how it improves patient outcomes and the usual approaches to model development. Health determinants, genomic and genetic data, electronic health records, wearable technology, and other data sources for predictive modeling are all fair game for papers. Every step of building a predictive model is addressed, including ethical considerations, data preparation, model selection, validation methodologies, and more. Patient participation, early disease detection, hospital resource management, personalized treatment planning, and other uses of predictive models are also included in the research. Evidence from case studies and other real-world instances shows that healthcare companies have used predictive analytics.

Chaudhary, Rohit et al., (2024) [9] this paper provides an in-depth review of the methodologies used to evaluate customer satisfaction data derived from airline tweets using machine learning algorithms. Airline firms who are looking to improve their service quality would do well to grasp the feelings conveyed in tweets, as social media has become an essential channel for consumer feedback. We take a look at how several machine learning approaches have been used to understand consumer sentiment and find commonalities in user feedback, such as topic modeling, predictive analytics, and sentiment analysis. The usefulness of several algorithms in extracting meaningful patterns from vast amounts of unstructured twitter data is highlighted in this paper, which carefully examines current research and case cases. These algorithms range from sophisticated deep learning techniques to more classic classifiers. The significance of proactive customer interaction techniques and real-time monitoring is highlighted as we go over the ramifications of these results for

airline management. Furthermore, we point out where the existing literature is lacking and propose new avenues for investigation, such as combining different types of data into a single dataset or creating more complex models for nuanced sentiment analysis. Researchers and practitioners in the airline sector who are interested in using machine learning to increase customer satisfaction might find a basic resource in this paper.

Ogrizović, Mihajlo et al., (2024) [10] In recent years, ML models have grown in popularity across a variety of applications, including computer vision and natural language processing, and they are almost always built on top of huge data. This is the intersection of software engineering, AI, and data science; an increasing number of products and apps come with integrated machine learning models. Operating in a real-world setting requires a system to address numerous challenges. These include designing for the possibility of incorrect model predictions, assuring safety and security in the face of such mistakes, determining which qualities are more important than a model's prediction accuracy, and identifying and measuring important quality requirements such as safety, scalability, explainability, fairness, privacy, and robustness. It is now essential to evaluate the capabilities and possible flaws of these models by testing them extensively. The purpose of enhancing and adapting current software testing methodologies is to find bugs in deep learning and machine learning models. This study presents a taxonomy, presents all solutions to the listed challenges in a consistent and methodologically sound manner, and draws inferences on potential patterns in future development. The paper's key contributions include a classification that adheres closely to the ML-pipeline structure, a clear definition of each team member's role within the pipeline, and a review of current trends and challenges in combining ML with big data analytics, with applications in both industry and education.

Udeh, Chioma et al., (2024) [11] Big Data Analytics is changing the way companies get insights from large and varied datasets in today's ever-changing world of business intelligence. This document summarizes the main points discussed in the extensive examination of Big Data Analytics and its effect on contemporary BI. Organizations may now tap into the full potential of their data assets thanks to Big Data Analytics, which signifies a revolutionary change in decision-making processes. The paper explores the many functions of Big Data Analytics, highlighting its

importance in customer-centric initiatives, operational efficiency, risk management, and strategic planning. When companies use predictive analytics to foresee changes in the market, their strategic planning improves dramatically. Better informed and future-oriented strategic initiatives are driven by decision-making that is aligned with organizational goals via the incorporation of analytics-derived insights. Integrating Big Data Analytics makes risk management more proactive, especially in preventing and detecting fraud. Reducing financial risks linked to fraudulent actions, the capacity to handle massive amounts of data in real-time increases awareness. Organizations may enhance their ability to anticipate and mitigate risks via scenario modeling. As analytics reveal inefficiencies in retail operations, supply networks, and industrial processes, operational optimization takes center stage. Retailers can stay nimble and sensitive to shifting consumer tastes and market conditions with the use of data analytics that allow real-time decision-making. Personalized marketing and predictive analytics in customer service boost customer-centric activities. Examining how Big Data Analytics helps businesses create customized experiences, this review seeks to increase consumer happiness and loyalty. The report summarizes Big Data Analytics' revolutionary path in contemporary BI, highlighting its function in easing the way through strategic complexity, reducing risk, improving operations, and making decisions with the customer in the focus. In this data-driven age, enterprises may get insights from the thorough evaluation that aims to help them leverage the transformational power of Big Data Analytics.

Mohammed Bachir, Mahdjoub et al., (2024) [12] Online stores can't survive without happy customers, therefore it's important to use CRM tactics that work to strengthen connections and boost performance. Although previous studies have shed light on the elements that impact customer satisfaction, such as payment methods, shipping times, and demographics, there are still many unknowns, especially when it comes to specific contexts like Olist, a well-known Brazilian online marketplace. This work seeks to fill these gaps by using state-of-the-art machine learning methods to a large dataset that includes 100,000 orders from 2016 to 2018. We use predictive methods such as logistic regression, decision trees, and random forests, as well as conduct thorough feature selection, to identify key factors that contribute to customer satisfaction. Our research provides practical advice for improving e-commerce success and highlights the critical role of operational efficiency in determining

consumer experiences. Not only does this study fill these research gaps, but it also provides practical techniques for improving customer happiness and directs future research in the field of e-commerce and CRM.

Fatima, Sheraz. (2024). [13] One game-changing way to improve patient outcomes and operational efficiency is by using machine learning (ML) in healthcare. Machine learning, by using big data analytics, enables the extraction of valuable insights from complex information; this, in turn, promotes customized treatment, early illness detection, and predictive healthcare metrics. Included in this paper's examination of machine learning's fundamental healthcare applications are illness diagnosis, therapy optimization, operational management, and patient risk categorization. Despite the promising results, there are still several challenges to using machine learning in healthcare. These include concerns about patient privacy, the intelligibility of models, biases in data, and the need for a robust infrastructure. To tackle these challenges, we need to work together across disciplines, establish appropriate regulatory frameworks, and improve machine learning algorithms so they are more scalable, equitable, and reliable. In order to completely use machine learning to revolutionize healthcare delivery, this research emphasizes the need to balance innovation with technological and ethical considerations.

Hanif, Atikah et al., (2023) [14] The aviation industry is only one of several that has seen the explosion of data produced by technological developments. These massive databases may be mined for useful information with the help of big data analytics. Twenty research publications were analyzed, with pertinent studies sourced from electronic sources including IEEE Xplore, Google Scholar, Web of Science, and ScienceDirect. Machine learning, "data mining," "predictive analytics," "big data analytics," and the "airlines industry" were among the search terms. The results show that Big Data Analytics (BDA) has several uses in the aviation sector, such as optimizing airline operations, improving service quality and customer happiness, enhancing safety and risk management, and maintaining aircraft. Considerations about data privacy and security, legal compliance, and data integration are some of the major obstacles to BDA adoption in the aviation sector that the analysis highlights. Ultimately, this review article is a great resource for learning about the benefits and difficulties of using BDA in the aviation sector. As so, it adds to what is

already known and lays the groundwork for further studies and real-world applications in this field.

Barik, Kousik et al., (2023) [15] Online shopping is always evolving and introducing new services in response to customer demand and technology improvements. An firm may improve its service and quality with the use of a strong customer satisfaction prediction model for privacy and trust platforms. A strategy to predicting customer satisfaction utilizing a blockchain-based framework that combines the MOL-PSOA with the MDNB-KNN was provided in this paper. We use a regression model to measure how different production variables affect consumer happiness. In comparison to earlier studies, the proposed method achieves better results when assessing accuracy (95%), precision (95%), memory (95%), and customer contentment (98%). Measuring customer happiness using a reliable platform may help with the development of the conceptual and practical distinctions influencing customers' buying decisions.

Rizwan Rashid, Muhammad et al., (2023) [16] With the rise of social media platforms like Twitter, Facebook, and Flickr, the majority of people now converse and share information online. The massive use of these social networking sites is constantly increasing the amount of free-text data. There is a great deal of unstructured data stored on these sites. The views of patients shared on social media greatly influence healthcare policy and play an important part in its advancement. We provide a machine learning strategy for healthcare feature optimum identification in this study. That strategy relies on an innovative synthetic technique. Moreover, we use an entropy-based method to label free-text remarks found in medical records as either favorable, negative, or neutral. Accuracy rates of 85%, 82.3%, 78.2%, and 87% between health care assessments were found in the experiments and evaluations. Our method, expert opinion, and patient interviews all show a weak correlation, as we found. We are able to provide a decent and accurate evaluation of the best healthcare facility for a patient by using machine learning methods to a degree of accuracy that shows we can do so. Based on social media reviews, our innovative framework can predict how patients will feel about their hospital stays. When compared to more conventional approaches, including polls and expert views, this new strategy clearly wins out.

Berros, Nisrine et al., (2023) [17] Data from public health services, epidemiology, public health administration, insurance companies, and traditional and non-traditional data sources like social media and online apps are just a few examples of the many sources of fresh imaging data that medicine is always producing. Many benefits have accrued to the healthcare industry as a whole from the use of big data, such as enhanced decision-support tools, more effective therapy, more tailored care, and more efficient clinical research procedures. Saving money and reusing it to boost efficiency and rationalization are two major benefits. The potential uses of big data in digital health are the focus of this article. We will go over the characteristics, quirks, and accessible tools for working with health data. Along with the previously mentioned technological and organizational obstacles, the most recent research pertaining to big data analysis in the health area is also highlighted. Lastly, we provide a broad approach for healthcare businesses considering big data analytics for adoption or use. This research provides healthcare organizations and institutions with a full grasp of big data analytics technology, including its possible uses, successful targeting, and predicted effects. It is useful for both those contemplating and already employing this technology.

Del Giorgio Solfa, Federico et al., (2023) [18] Machine learning and big data analytics help medical professionals choose the best course of therapy, allocate resources efficiently, and create individualized prescription regimens. To guarantee that algorithmic recommendations are impartial and fair, it is necessary to think about ethical issues connected to prejudice and data privacy. There will be a dramatic shift in the medical industry as a result of machine learning and big data analytics, opening up new possibilities for reform and better patient outcomes as these technologies develop further. Using an online poll that included healthcare experts, patient evaluations, and clinical staff, this study aimed to empirically assess the patient's results. The structural model was predicted from the data using SmartPLS 4.0. The results showed that machine learning has a direct and beneficial effect on healthcare performance and patient outcomes when applied to big data analytics. In addition, it's clear that this may result in tailored treatment programs, prompt interventions, and better health results for patients. Healthcare organizations may also benefit from big data analytics in terms of cost reduction, operational efficiency, and optimal resource allocation. Big data analytics is a promising field that deserves more funding and

attention because of the predicted increasing influence it will have on healthcare performance and patient outcomes.

Rangineni, Sandeep et al., (2023) [19] Academic investigations on data quality and its far-reaching effects on analytics. Data quality encompasses several aspects, including completeness, timeliness, dependability, correctness, and completeness. The immaterial starts with explaining these characteristics. In order to fully use data analytics methods and approaches, it is crucial to understand and handle these aspects. The abstract continues by discussing the difficulties of guaranteeing data quality, such as problems with data integration, data purification, and the ever-changing character of data sources. Also included in this abstract are the methods and strategies used to improve the data quality. The importance of data profiling, data cleaning, and data standardization in detecting and fixing data discrepancies is emphasized. Case examples from the real world highlight the critical link between high-quality data and the efficacy of data analytics approaches. Investment in data quality projects yields measurable advantages, as shown in these case studies. These benefits include better decision-making, higher customer satisfaction, and simplified operational operations. Furthermore, the abstract delves into the consequences of inaccurate data, including misguided company strategies and inaccurate prediction models, highlighting the potential financial and reputational hazards linked to inadequate data. It advocates for a preventative approach wherein companies use robust data governance frameworks, state-of-the-art technology, and skilled personnel to ensure the data's consistent quality. Doing so will allow companies to tap into data analytics' full potential, which will lead to innovation, competitive advantage, and long-term success.

Alzaabi, Omar et al., (2023) [20] Big data allows for the rapid and accurate analysis of data that was either inaccessible or underused before. Because of this, analysts, researchers, and business users can make better decisions. Text analysis, machine learning, predictive analytics, data mining, statistics, and natural language processing are all examples of advanced analytics technologies that may help organizations find new insights, whether they utilize them in conjunction with their existing data or on their own. There is a plethora of database management systems, strategies, and metrics that may be used to leverage big data in every facet of life, and the majority of national, regional, and international agencies are now tackling big

data. However, there are still significant challenges in dealing with this data and making the most of it. Using the healthcare industry in the United Arab Emirates as an example, this article explores how big data analytics affects project risk management. A literature study investigating the impact of data analytics implementation on business, technology, industry, and society as a whole was produced by an exhaustive search of over twenty references. Data analytics' function in healthcare data analysis and management, its advantages in risk prediction and healthcare outcome improvement, and future insights have been the subject of an interview-based study technique. The findings of this study highlight the significance of data analytics in several industries, including healthcare and business, as a kind of digital transformation. One clear takeaway from this study is the widespread use of business intelligence and data analytics in the healthcare sector of the United Arab Emirates. This has opened up many doors for new prospects and highlighted the need of positive change in this industry.

Gkioka, Georgia et al (2023) [21] The amount of user-generated content (UGC) expressing opinions on the quality of services offered has grown at an exponential pace. The use of data-driven methodologies to comprehensively assess service quality has received less attention in the research community. Applying Data Analytics (DA) methods to user-generated content (UGC) has been shown to be an effective and promising way to gauge service quality in a number of studies. Finding research gaps and suggesting future paths are the goals of this study, which use data analytics tools to assess service quality across industries. The purpose of this research is to assess service quality across different industries by conducting a comprehensive literature review (SLR) of data analytics (DA) methods. The data, methods, and assessment strategies employed in these investigations are the main points of this article. Findings from the research include a new taxonomy for Data Analytics approaches to service quality measurement, identification of the most popular data sources, and industry-specific insights on the methodologies and data sources employed. To wrap up, the study points out where the literature is lacking and suggests ways forward for research that will hopefully help both academics and practitioners use DA as a supplement to more conventional survey-based techniques of service quality evaluation.

Janani, S.R. et al., (2023) [22] This article details the development of a healthcare analytics and monitoring system that utilizes data collected from several IoMT devices to make predictions about patients' cardiovascular risk using machine learning. Following a feature selection procedure in a random forest, the study shifts to a fuzzy logic classifier to forecast CVD. A 7-step simulation is conducted to ascertain the relative efficacy of the proposed machine learning-based data analytics methodology compared to other methods. The results demonstrate that the suggested method achieves better CVD classification accuracy, recall, and F1-score than the current state-of-the-art.

Lee, Seon Hong et al., (2022) [23] Many people rely on metaverse technology for things like education, tour systems, and mobile application services on a daily basis. Because they are accessible from anywhere, mobile metaverse apps in particular are seeing a rise in user numbers. Analyzing online reviews that convey consumer happiness is vital for improving the service. In order to investigate user satisfaction with little survey data, a number of earlier research have used conventional approaches like the SEM and the TAM. Analyzing mobile app users may not be a good fit for these methodologies. Several academics have found ways around this issue by analyzing user experiences via star ratings and online reviews. Having said that, there are instances when the tone of the review does not match the star rating. This inconsistency hinders machine learning's efficiency. A lexicon-based sentiment classifier called VADER is proposed to help with the differences. The present research endeavors to construct an improved sentiment classifier using VADER as a machine learning framework. Using five sentiment classifiers—Naïve Bayes, KNN, Logistic Regression, Light Gradient Boosting Machine (LightGBM), and Categorical boosting algorithm (Catboost)—along with three embedding methods—Bag-of-Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), and Word2Vec—constituted this study. With the exception of one classifier (Logistic Regression with Word2Vec), the findings demonstrate that classifiers using VADER perform better than those without VADER. With an accuracy of 88.68%, LightGBM with TF-IDF outperforms all other models.

Esposito, Francesco et al., (2022) [24] Globalization and fast innovation have opened up a world of possibilities for businesses and consumers alike. The quantity of

products created has skyrocketed due to the obvious rise in production and sourcing on a global scale brought about by competitive forces. Finding out how much interest there is in supply chain analytics for real-time business information is the goal of this research. An analysis and explanation of BI's advantages and challenges, as well as the methodology and strategy, are presented in the article. In order to aid service-oriented firms in their pursuit of customer retention and loyalty, the article focuses on the need to reconsider the conventional BI concept of integrating and consolidating a company's data. Supply chain analytics may be made more effective and efficient with the use of a business intelligence plan, which is essential for a company's competitive edge. This article helps readers better grasp the challenges of implementing BI tools into supply chains.

Helae, Maryam et al., (2022) [25] Using the random forest (RF) model of machine learning, this research aims to analyze patient evaluations for sentiment in order to determine how satisfied they are with various medications. Also, by using the long short-term memory, we may determine which medications are most effective for certain diseases by analyzing patient feedback. Lastly, the writers use their reviews to make predictions about the patients' medical problems using support vector machine and RF classifiers. The pharmaceutical and medical outcomes will noticeably improve if we know the patients' medical conditions and levels of satisfaction.

Batko, Kornelia et al., (2022) [26] Now that Big Data Analytics (BDA) is here, medical professionals may use cutting-edge tech to better track their patients' health and treat them. The primary goal of this piece is to investigate possible healthcare uses of Big Data Analytics. This inquiry is built around a literature analysis and the presentation of important results from empirical investigations on the application of Big Data Analytics in healthcare settings. The questionnaire-based direct survey drew from 217 healthcare establishments in Poland. Medical facilities in Poland are moving towards data-based healthcare, utilizing both structured and unstructured data and extending analytics to administrative, business, and clinical domains. This shift is backed by literature studies that highlight the many advantages of Big Data Analytics. The results of the research supported earlier hypotheses that healthcare companies are dealing with structured and unstructured data simultaneously. Information gathered by devices and sensors is different from data

stored in databases or processed via transactions. Unstructured data found in emails and documents is another sort of data. They use analytics in their clinical, administrative, and business work, but they don't depend as much on social media data. Statistics unquestionably have a role in healthcare facility decision-making. The research found that data-based healthcare offers several benefits, and medical institutions are increasingly embracing it. The literature has already addressed this tendency.

Ikegwu, Anayo et al., (2022) [27] Modern industrial operations, corporate intelligence, and the way industrial revolutions are seen are being transformed by the study of big data analytics (BDA) techniques for data-driven industries. Due to its singularity, big data and BDA have sparked a flurry of new research requests aimed at finding answers to problems with data processing, storage, visualization, and creation. Researchers and practitioners alike are woefully uninformed when it comes to the best ways to manage big data formats, as well as the information and BDA tools necessary to extract insight from massive amounts of industrial data. Regardless, there have been a number of current proposals for academic studies and research into big data analytics as a means to enhance industrial performance. Methods, data sources, applications, important obstacles, and suitable solutions have not yet been thoroughly reviewed, compared, and evaluated rigorously. The following contributions are made by this study to address this gap: takes a comprehensive look at the latest developments in BDA methodology, tools, and trends, as well as their advantages and disadvantages. Find and talk about data sources and practical uses where BDA might make a difference. In addition to outlining problems with and potential solutions to BDA, this report also identifies areas of future research that should be further investigated. Researchers, industry professionals, big data providers, and governments working on BDA might benefit from this study's informative advice about the problems with existing BDA approaches and how to fix them.

Verma, Navneet et al., (2022) [28] Globally, health monitoring is a focus of sustainable development, and diabetes mellitus is a leading cause of health crises due to its prevalence. Presently, a very effective method for tracking and forecasting diabetes mellitus is offered by combining IoT and ML technologies. We have presented a model in this article that utilizes the LoRa protocol of the Internet of

Things (IoT) and the hybrid enhanced adaptive data rate (HEADR) algorithm to collect real-time data from patients. Moreover, the data acquired via the LoRa protocol is used to make machine learning predictions about the severity of diabetes using categorization algorithms. Using throughput and packet collision metrics, the Contiki Cooja simulator assesses the LoRa protocol's performance. Using the Python programming language, the suggested model employs various machine learning classifiers to forecast the occurrence of diabetes.

Aattouchi, Issam et al., (2021) [29] "Tweets" are 140 characters in length and are sent and viewed by users of the microblogging website Twitter. Twitter is quickly becoming a popular arena for medical research, with many unstructured and free-text tweets pertaining to healthcare being posted on the platform. Among the many forms of data mining, sentiment analysis is useful for predicting how NLP will use personality traits to analyze sentiment. Computational linguistics is a field that uses text analysis to infer and study user knowledge about the internet, social media, and related topics. By quantifying societal sentiments towards certain products, persons, or ideas, the examined data reveals the contextual duality of knowledge. Many industries, including healthcare, make use of sentiment analysis. Social media and websites that rate medical issues are only two examples of the vast amounts of unstructured healthcare information that is accessible online. Among sentiment analysis' many uses is the enhancement of health care quality and the attainment of optimal patient outcomes via the use of medical data. Methods for sentiment analysis in the medical domain are the subject of this review article.

Salman, Omar et al., (2021) [30] To accommodate everyone, save more lives, and manage medical resources efficiently, triage and categorizing patients into multi-emergency levels is necessary due to the dramatically rising numbers of patients. When patients use telemedicine systems and live a long distance from the hospital, triaging and prioritizing their care becomes an even bigger challenge. Researchers are taking use of machine learning, a powerful technology, to tackle this difficulty. As a result, there has to be a thorough examination of the topic of using machine learning to E-triage and patient prioritization. (1) This research will survey the existing literature and conduct an extensive analysis of machine learning's functions in E-triage and telemedicine's patient prioritization systems in order to improve the

efficiency and effectiveness of healthcare delivery. (2) draw attention to the achievement of machine learning objectives in telemedicine for distant healthcare systems by analyzing algorithms, medical input data, and output outcomes. (3) show how machine learning objectives relate to electronic triage procedures, with a focus on the following: relevant illnesses, output outcomes, input medical characteristics, and triage levels. (4) We use the results of our research to categorize and suggest a hybrid taxonomy of telemedicine frameworks and machine learning methods. (5) provide academic and industry stakeholders with lists of reasons for doing research into telemedicine and remote healthcare applications, as well as open research issues and suggestions for future intelligent work. Approach: We conduct a comprehensive literature study on E-triage and remote prioritization systems that use sensors and machine learning techniques. In order to find publications published between January 2012 and the present, we have searched all relevant terms in the following databases: ScienceDirect, IEEE Xplore, Web of Science, PubMed, and Medline. The outcomes are: We provide a novel approach to telemedicine taxonomy and machine learning algorithms for crossover matching. This research uses a crossover-taxonomy to determine the nature of the connection between ML algorithms and their corresponding telemedicine categories. The use of machine learning has a significant influence on the proposal of synchronous (online, in real time) and asynchronous (store-and-forward, offline) telemedicine architectures. Further information is provided in the form of a list of machine learning algorithms, performance measurements, input data, and output outcomes. The presentation is also organized and classified by open research questions, machine learning's benefits, and potential new areas of study that may be explored to bring together disparate fields in a more harmonious way. Point of view: Recent studies on E-triage and prioritizing systems using telemedicine architecture and machine learning approaches are included in this area. This approach has the potential to promote research on healthcare system modernization and the effective application of ML and AI. When it comes to tracking the well-being of the growing number of people across the world who suffer from diabetes, hypertension, cardiovascular disease, and other chronic illnesses, E-triage and prioritizing systems powered by machine learning algorithms could prove to be quite useful. To sum up, telemedicine's e-triage and prioritization systems using machine learning algorithms have distinct but complementary domains of study. Consequently, we provide an in-depth evaluation to emphasize the advantages of

ongoing research in multidisciplinary endeavors related to healthcare services, AI, and ML.

Habehh, Hafsa et al., (2021) [31] Predicting health crises, identifying disease populations, and predicting disease states and immunological responses are just a few domains that have reaped the benefits of recent advancements in AI and ML technology. Despite ongoing skepticism about the practicality and accuracy of ML-based approaches in healthcare settings, their use is on the rise. Through the use of concrete examples, we provide a brief overview of supervised, unsupervised, and reinforcement learning, as well as other machine learning-based learning algorithms and approaches. Second, we discuss the many applications of ML in healthcare, including as radiography, neuroimaging, genomics, and electronic health records. In addition to discussing potential future applications of ML, we briefly touch on some of the risks and challenges that come with using it in healthcare, such as concerns about system privacy and ethical considerations.

Ahmed, Imran et al., (2021) [32] Many diverse areas of life make use of smart sensors and IoT, including commerce, healthcare, transportation, communication, and surveillance. Big data analytics comes into play when these IoT devices and sensors generate vast amounts of data that, if analyzed, may be valuable and useful for healthcare companies. A new coronavirus pandemic (COVID-19) has just broken out, posing a grave danger to people's lives, livelihoods, social connections, and even global relations. The IoT and big data have been important in this battle against the epidemic. Possible uses include efficiently gathering large amounts of data, visualizing pandemic information, assessing the effectiveness of measures to prevent and control the spread of COVID-19, keeping tabs on confirmed cases, monitoring prevention levels, and so on. We provide a health monitoring platform for COVID-19 analysis and prediction in this research. The IoT and big data analytics are used by the framework. We use a unique illness actual data set to do descriptive, diagnostic, predictive, and prescriptive analyses utilizing big data analytics. Our emphasis is on various pandemic symptoms. This research primarily contributes by analyzing and forecasting a new disease using Big Data Analytics in conjunction with the Internet of Things. A neural network-based pandemic detection and forecasting algorithm has been created to assist medical staff. To predict the spread of pandemics, we use neural

networks and evaluate their efficacy in comparison to existing ML systems. A 99% accuracy rate puts the neural network ahead of the competitors, according to the data.

Saini, Mukesh et al., (2021) [33] Improved patient profiles and prediction models made possible by healthcare data are allowing doctors to better anticipate, diagnose, and treat a wide range of ailments. Furthermore, data pools have been developed via collaborations and partnerships between healthcare groups and academics; these pools may be used to create more accurate customized healthcare models. New approaches to healthcare are emerging as a result of the widespread use of AI and ML. With the use of statistical tools and algorithms, researchers and medical practitioners now have real-time access to all clinical case studies and white papers conducted on genetic disorders. This not only aids in the efficient detection of healthcare fraud but also speeds up the development of more precisely targeted vaccines. Additionally, it helps find more cost-effective ways to analyze diseases and treat patients effectively. Researchers and medical professionals can gain a better understanding of a patient's condition and its full scope with the help of this proposed mechanism that reads their clinical records and radiology data. This will not only speed up the process of finding a cure and treatment options, but it will also aid in the healing of patients who are in need of healthcare.

Devesh, U et al., (2021) [34] Innovative medicines, expanded physician responsibilities, new laws, payment models, and healthcare IT are all likely to make health care more complicated. Big data is being created as a result of the rapid evolution of technology, which is derived from a variety of resources ranging from very basic to quite sophisticated. In order to build models of pharmacy practice, Big data is here to stay, and everyone involved in the healthcare system has to understand and take use of it. The ever-growing volume of organized and unstructured biological and biomedical data acquired from many sources, such as hospitals, laboratories, and pharmaceutical companies, necessitates the use of data mining and machine learning algorithms for appropriate assessment. Providing an opinion on how healthcare systems might benefit from big data is the main goal of this essay. Because health care data is always growing in volume and complexity, big data has the potential to improve patient-centered pharmaceutical service. Diverse data analysis methodologies can manage data pertaining to electronic health records and medical imaging,

according to current study that includes creating and using a specific healthcare framework. With the support of strong integration of healthcare and biological data, contemporary healthcare providers may likely revolutionize individualized medicine and medical therapy.

Vargas-Calderón, Vladimir et al., (2021) [35] In order to enhance hotel services and boost the quality of service perception, it is vital to get firsthand knowledge about customer preferences, and online hospitality platforms are providing just that. Automatically extracting the most important components of service quality for hospitality clients may be done using customer reviews. This study presents a methodology for evaluating service quality in the hotel industry using real-world customer evaluations and NLP/ML techniques. The suggested system can figure out what matters most to hotel guests in terms of service quality automatically. Reviews of hotels in Bogotá and Madrid are immediately removed from Booking.com. By using FastText and Latent Dirichlet Allocation, which enable the representation of text reviews as vectors, semantic information to be derived. To make sense of and make sense of massive volumes of user evaluations, we use a dimensionality reduction approach. In order to evaluate the service's quality both numerically and subjectively, visualizations of its most crucial features are created. The results demonstrate that big customer review datasets may be automatically mined for the most important characteristics of service quality as evaluated by consumers. Managers in the hotel industry should utilize these results to have a deeper understanding of their customers and provide better service overall.

Hosseinzadeh, Mehdi et al., (2021) [36] The generation of digital data is always increasing as new technologies are developed. Therefore, data management solutions are required for dealing with the massive datasets. Because of issues like noise, missing values, inconsistencies, and so on, the data collected from many sources—such sensor networks, social media, company transactions, etc.—is fundamentally ambiguous, lowering the quality of big data analytics. Identifying and fixing corrupted data, also known as data cleaning, is a major obstacle in this setting, and several approaches have been proposed to address this problem. Nevertheless, as far as we are aware, no extensive analysis of data cleaning methods for big data analytics has been conducted. Therefore, this survey conducts a thorough and

methodical investigation of the current state of the art procedures for cleaning huge data. As a result, we classify these methods into five groups: those based on machine learning, samples, experts, rules, and frameworks. Each category has a number of articles that have been evaluated. In addition, the study compares the selected data cleaning procedures based on their scalability, efficiency, accuracy, and usability, and it identifies their benefits and drawbacks. It also examines the relevant criteria. Lastly, several recommendations for future research are offered to enhance the massive data purification processes.

Lee, Hong Joo et al., (2021) [37] Due to the enormous amount and noisy character of big data, very few studies have examined social media attitudes toward public services, despite the increasing need and potential for social media to aid in government decision-making. This research proposes a way to systematically assign tweets into each of the SERVQUAL dimensions in order to discover attitudes and monitor policy makers' perceptions of the service quality of healthcare services. The strategy is based on design science. First, the technique uses clustering and recurrent latent Dirichlet allocation (LDA) to find more trustworthy topic sets; second, it uses a preexisting notion of service quality to categorize tweets based on their themes. To prove its usefulness, the approach is applied to tweets about the UK's National Health Service (NHS) quality. For each SERVQUAL dimension, we determined keywords and assessed how the public views the quality of healthcare services. Additionally, the suggested strategy is confirmed to be useful by comparing the social perceptions gained from tweets with conventional survey results on the same service quality. Academically, the technique is valuable as a unique approach that integrates text mining with the theoretically sound quality framework, SERVQUAL, and practically, it is useful as a supplement to the more costly national scale surveys.

Rahim, Afiz Izzudin et al., (2021) [38] A new channel for hospitals and patients to communicate about treatment quality is social media. Automated sentiment and SERVQUAL dimension classifications were achieved by the use of machine learning in our sentiment analyzer and classifier. In order to establish our objectives, we used logistic regression analysis. We used a machine learning sentiment analyzer to sift through 1852 reviews; it identified 72.1% favorable and 27.9% negative. With the use of our machine learning SERVQUAL classifier, we were able to categorize

240 reviews as concrete, 1257 as reliable, 125 as reacting, 356 as providing assurance, and 1,174 as demonstrating empathy. All SERVQUAL aspects, with the exception of Machine learning algorithms enabled by Facebook reviews provide real-time data that conventional hospital quality evaluations can overlook. All parties involved in healthcare may also gain from the hitherto unrealized value of online patient reviews as an indicator of quality. We found that Facebook reviews are a useful supplement to other methods for gauging hospital service quality in Malaysia, and our findings corroborate those of previous research.

Nawab, Khalidet al., (2020) [39] As a result of hospitals' financial incentives that depend in part on patients' satisfaction with their treatment, gathering and analyzing patient feedback is a top priority. Here, we show how to utilize NLP to glean useful information from Press Ganey survey responses provided by patients. Programmatically standardizing textual data using NLP tools was the first step. Among these tasks was the elimination of unnecessary words, the change to lowercase, and the correction of spelling errors. After that, we made charts out of numerical data for each category according to the care and emotion component. For our in-depth analysis, we focused on the care component categories with the highest number of unfavorable responses. We created tables of the most common words, adjectives, and bigrams using natural language processing. To further understand the elements that contribute to unfavorable patient feedback, comments that included common terms or combinations were manually further studied. After sorting the reviews by positivity and negativity, we trained a neural network to analyze the sentiment of the phrases derived from the splits. Patients clearly placed a high value on physicians and nurses, as the majority of the remarks pertained to these healthcare professionals. Out of all the categories, "room," "discharge," and "tests and treatments" garnered the most unfavorable responses. Words, adjectives, and pairs of words that appeared often were then entered into a database. Our research shows that several factors contribute to patients' bad experiences, including room temperature, cleanliness, and noise levels, as well as delays in discharge paperwork, contradictory information on the discharge plan, multiple blood samples, and excessive needle jabs. Results derived only from numerical data did not reveal any of this. Natural language processing (NLP) is a potent technique for rapidly processing massive volumes of

patient data because it can acquire insight from raw textual feedback and extract relevant information.

Shah, Adnan et al., (2020) [40] A number of studies have used textual contents to analyze patients' opinions. In order to determine whether healthcare service delivery is of high or poor quality, this research introduces a new multimodal technique. Unlike previous research, we take into account both the unique textual and pictorial contents of the Yelp.com platform. This is more difficult than before because of feature extraction. Using both visual and verbal clues, as well as a combination of the two, we assess how well deep learning algorithms and the baseline perform. Adding new features improves the models' accuracy by 15% across both text and visual signals, according to the experimental data. In addition, compared to only looking at the contents, the classification accuracy increases by a minimum of 12.64% when text and visual contents are fused. According to the results of the preliminary trials, the deep learning model performed better than any of the other methods. The suggested model was able to considerably improve the classifier's performance when compared to several other state-of-the-art approaches in the biomedical area, suggesting that the methodology is effective and appropriate.

Zaman, Nohel et al., (2020) [41] Administrators at healthcare facilities constantly evaluate a plethora of aspects with the goal of increasing patient happiness, which is strongly correlated with patients' preferences in healthcare providers and the quality of their medical results. Then, to get a feel for how satisfied customers are with the service as a whole, we apply regression modeling to find out which service characteristics correlate best with star ratings. Results show that patients are more likely to score things like wait times, treatment efficacy, communication, diagnostic quality, environmental cleanliness, and cost when these factors are mentioned. Lastly, we automate the process of detecting customer references of these service qualities by building text analytics. From this, we extract a list of "smoke terms," or phrases that appear often in postings referring to certain features of services. Our findings suggest that standard sentiment analysis is not very useful as these words are often not emotional. For managers, this data is useful for identifying which areas are most in need of development. Health care professionals and policymakers may greatly benefit

from smoking words for each service feature when it comes to quickly monitoring issues and making adjustments to policies or resources to enhance service.

Wassouf, Wissam et al., (2020) [42] In today's commercial world, client behavior is more important. Telecom operators recognize this and want to improve their market share by focusing on customer profitability, customer loyalty, and churn. A new era of CRM tactics was ushered in with the advent of big data principles. The Syriatel firm provided the dataset that was used to develop and test this approach. The dataset comprises around 127 million records for training and testing. First, the new technique was used to categorize clients based on time-frequency-monetary factors. TFM (TFM locations: The sum of all phone conversations and online sessions during a certain time period is represented by the symbol "time" (T). Frequency (F): make heavy usage of services over a certain time frame. "Monetary" means the amount of money spent during a certain time frame. and each group's degree of devotion was specified. Second, we used customer demographics (age, gender, and shared services included) and loyalty level descriptors as categories to choose the most useful behavioral attributes based on customer behavior. Thirdly, several classification techniques were used to develop prediction models that could categorize new users according to their loyalty. These algorithms were based on the descriptors and attributes that were selected. At last, the models were assessed using various criteria in order to generate the rules for loyalty prediction. In order to provide them with the most relevant offers and services, we first analyzed these criteria to determine the loyalty reasons at each level.

Kasten, Joseph. (2020).[43] Especially in clinical settings, the healthcare sector is increasingly making use of big data-related technologies including data analytics, the internet of things, and machine learning. But, similar to any other kind of company, healthcare facilities are also required to carry out all administrative tasks. Consequently, healthcare has started to leverage these similar technologies to better understand their internal operations, which allows them to run a more effective operation and make better choices, much like many other businesses. In order to characterize the existing literature and to identify the key topics and goals of the research community and the healthcare business overall, this study employs a systematic literature review methodology. In order to provide the groundwork for the

future research possibilities mentioned in the article, this study contributes by offering a thorough evaluation of the current literature.

Lopez, Fernando et al., (2020) [44] A machine learning and data integration–based intelligent big data analytics platform is already under development at the Keralty business. In this article, we'll go over the new platform and how it will help the organization tackle healthcare's future problems, including value-based care, population health management, and more. Better healthcare outcomes, enhanced clinical operations, decreased healthcare expenditures, and accurate medical information collection are some of the advantages of using this new data platform for population and community health. By using the platform's massive standardized datasets, the authors' machine learning algorithms may enhance public health initiatives, diagnostics, and clinical decision support tools. The platform incorporates data from various sources, including public health platforms, mobile data, social media, clinical web portals, and Radiology Information Systems (RIS), Laboratory Information Systems (LIS), Electronic Health Records (EHR), and Hospital Information Systems (HIS). The storage, retrieval, processing, and transformation of this enormous data set is accomplished via the integration of big data approaches.

Zunic, Anastazia et al., (2019) [45] Automatically categorizing the sentiment conveyed in free text is the goal of sentiment analysis (SA), a branch of natural language processing. Focusing on user-generated material rather than medical experts' viewpoints allowed us to better understand the impact on people's health and wellbeing. Methods: Systematic review protocols served as the basis for our approach. We searched the MEDLINE database for relevant articles in January 2019 using the multi-faceted interface PubMed. Out of 86 research that met our criteria, we extracted information on the datasets examined, discourse subjects, data producers, downstream applications, methods used, and assessment results. Findings: Online shopping and social media sites accounted for the bulk of the data. The two main goals of most internet chats are knowledge sharing and offering social support. These groups often coalesce around very serious and long-lasting medical issues. Drugs, immunizations, surgeries, orthodontics, doctors, and medical care as a whole are just some of the topics covered. In this study, we looked at several kinds of spontaneously created tales and found that the writers may be in one of five health-related roles:

sufferer, addict, patient, caregiver, or suicide victim. Only four research (out of eighty- six total) really recorded these demographic details. To carry out SA, a variety of techniques were used. Support vector machines, decision trees, naïve Bayesian learning, logistic regression, and adaptive boosting were the most popular options. Despite the prevalence of deep learning in SA research, only a single paper actually used it. When compared to other domains' state-of-the-art, the performance falls short with an average F-score below 60%. When looking at South African context, the health and wellbeing domain was determined to be resource deficient, with very few publicly available domain-specific lexica and corpora. Results from SA in the field of health and wellbeing are behind those in other areas, according to the conclusions. We don't yet know whether this is due to algorithmic choices, small training datasets, a lack of domain-specific sentiment lexica, or inherent disparities across domains and sublanguages.

Thangarasu, Gunasekar et al., (2019) [46] It has given healthcare organizations the means to collect, organize, and analyze both structured and unstructured data. The healthcare sector has recently seen a rise in the use of big data analytics to aid in illness diagnosis and treatment. But, owing to some basic issues intrinsic to the big data paradigm, the healthcare sector is still sluggish in its acceptance and research development of big data analysis. This study offers a fresh approach to large data analytics using Apache Spark, tackling these issues while concentrating on the future of medical research. The healthcare industry's ability to provide treatment will be enhanced by the suggested method. Improved patient care and more efficient medical procedures are both possible outcomes of big data analytics' ongoing evaluation of clinical data.

Abidi, Syed et al., (2019) [47] A large amount of diverse data is produced by the healthcare system, which is a living system. Intelligent and interactive analysis of health data is becoming more important for health system management as healthcare organizations transition to value-based systems. This is particularly true for optimizing resources while enhancing service quality and health outcomes. Innovations in AI and big data are reshaping health data analytics with fresh ideas and smart approaches. Here, we put health data analytics and data collected from health records in the perspective of current developments in AI and big data. In order to

determine "how big" health data is, we analyze its characteristics according to the big data criteria. The next step is to describe how data analytics techniques based on artificial intelligence function. Then, we'll talk about "what insights" that may be gained from a variety of health data analytics approaches to enhance healthcare innovation, knowledge discovery, health outcomes, and system management.

C S, Kanimozhi et al., (2018) [48] Any item may be the subject of people's thoughts, comments, recommendations, or ideas. There are a lot of places where people may voice their opinions, such reviews, blogs, Facebook, Twitter, etc. In today's rapidly evolving world, individuals increasingly rely on online platforms like blogs, social media, and shopping websites to gather information before making important life choices like purchasing a product or casting a ballot for a candidate. Few studies have focused on health and medical domains within opinion mining, as opposed to more broad areas like consumer electronics, entertainment, and restaurant evaluations. Those struggling with substance abuse frequently turn to the internet for support, since they may not be able to find it among their loved ones [1]. This study makes use of an opinion mining technique that attempts to forecast the degree of drug satisfaction amongst other patients who have previously felt the effects of a medicine. The satisfaction level of patients is predicted by analyzing the sentiment of each review and then using the classification algorithms Naïve Bayes and k-nearest neighbor. In this research, we have used two medications for the same condition. We compare the datasets' polarities to choose the most suitable one.

Koren, Gideon et al., (2018) [49] Despite the availability of excellent medications, Uncontrolled hypertension continues to be prevalent. Systematic reviews and randomised controlled trials provide the basis of treatment plans. When applied to large datasets, machine learning the researchers aimed to discover anything new about hypertension medicine. Our goal was to identify the variables that influence the success of hypertension medication treatment by analyzing data from a large patient population using two machine learning tools: decision trees and neural networks. Additionally, we discovered drugs that are often used concurrently yet are not believed to lower blood pressure (BP). Higher blood pressure at baseline is a predictor of worse outcomes. Among the pharmacological options and combinations, beta blocker treatment seems to be more generally effective, albeit this is not reflected in

current guidelines. Patients with hypertension sometimes take many drugs at once, including statins and proton pump inhibitors (PPIs), which significantly increases the therapy efficacy rate. In conclusion, a new strategy for the effective discovery of antihypertensive medicines and the identification of novel applications for current medications is offered by machine learning in the context of big data. Machine learning applied to a diverse and extensive set of big data has shown promise for an underutilized class of drugs, beta blockers, and we recommend that other methods confirm and validate these findings. These methods differ from the much more limited criteria used in randomized clinical trials (RCTs). Supporting our big data results are the hitherto unrecognized benefits of proton pump inhibitors and statins on lowering blood pressure in early clinical observations.

Wang, Lidong (2016) [50] One kind of artificial intelligence, machine learning, may discover previously unknown data and make intelligent decisions. The impact of Big Data on scientific advancements and economic growth is substantial. Machine learning methods, Big Data infrastructure, and various ML uses within Big Data are all summarized in this article. Several difficulties arise when combining machine learning with Big Data, and we go over them here. It also showcases new approaches and advancements in machine learning for Big Data. The official name of the journal is "International Journal of Mathematical, Engineering and Management Sciences" (2014).

Archenaa, J. et al., (2015) [51] This article offers some suggestions for making better use of the data produced by healthcare and government agencies. These agencies provide a great deal of data that is not uniform in nature. But these data were meaningless without the right data analytics techniques. The use of Hadoop for Big Data Analytics is crucial for making sense of massive amounts of data in real time and foreseeing potential crises. It explains how the government and healthcare industries are using big data.

Na, Jin-Cheon et al., (2015) [52] Since medicine review websites are relatively unexplored in comparison to more generic domains like product reviews, this work fills that need by developing an efficient approach for sentiment analysis of user-generated material on these sites. Sentences may comprise numerous clauses covering various features of drugs, hence an algorithm for clause-level sentiment analysis is

built. This method takes a purely linguistic approach to determining whether a clause is positive, negative, or neutral in sentiment based on the previous sentiment scores given to words within the clause, with the grammatical relations and semantic annotations (like disorder terms) of those words taken into account. The usefulness of the suggested technique was shown by experimental results using 2,700 clauses. It outperformed the baseline approaches utilizing a machine learning approach. By analyzing mistakes, we were able to find and talk about many difficult problems. Patients, medication manufacturers, and healthcare providers may all benefit from the suggested sentiment analysis technique, which provides relevant public opinion summaries. Incorporating domain information about medication reviews into the sentiment analysis algorithm allows for more precise analysis, since sentiment analysis is domain specific. Specifically, the Unified Medical Language System (UMLS) Semantic Network makes use of MetaMap to associate different health and medical words (such names of diseases and drugs) with certain semantic kinds.

Like Oxygen et al., (2015) [53] Nowadays, data is everywhere. The amount of data that we collect and consume is growing rapidly in the digital age. More and more people are using new technologies and social media, which produces massive amounts of data that, when examined correctly, may provide wonderful insights. Big data refers to very massive datasets that are too huge to be stored in conventional databases. In order to improve decision-making and results, organizations must handle and analyze massive data. Big data analytics, then, is all the rage these days. Big data analytics has great promise for enhancing patient care and providing clinical decision support in the healthcare industry. In this article, we take a look at the history and different approaches to healthcare big data analytics. This study also discusses the pros and cons of big data analytics and provides an in-depth analysis of the several platforms and algorithms used for this kind of analysis. Finally, this study concludes with a look forward at potential obstacles and solutions.

Tan, Sharon et al., (2015) [54] This editorial is a part of the "Big Data and Analytics in Healthcare" Focus Theme of the Methods of Information in Medicine. The healthcare business is seeing a meteoric rise in the quantity of data being created. The availability of healthcare data, sometimes known as "big data," has sparked a surge of interest in finding ways to use it to better health outcomes while cutting

costs. But there are significant difficulties in handling and interpreting big data in healthcare due to the specific characteristics of healthcare data. New ways of thinking about and talking about these problems are what this Focus Theme is all about. Approaches that potentially provide improved predictions from healthcare data are detailed, including efficient techniques of processing huge clinical datasets and predictive models.

Hopper, Anthony et al., (2015) [55] In order to quantify text-based information found on the internet, this research aims to examine the utility of sentiment analysis and time-to-next-complaint methodologies. More importantly, the authors show how managers can employ time-to-next-complaint approaches to arrange data obtained from sentiment analysis into meaningful information that physicians and other staff can share. The authors reviewed patient feedback for a subset of Virginia gynecologists using sentiment analysis as their technique and strategy. In order to make sense of this data, the writers used time-to-next-complaint procedures among other ways. Conclusions - The authors showed that healthcare administrators would benefit from sentiment analysis and time-to-next-complaint methodologies if they wanted to turn web-based material into valuable, quantitative data. Implications and limits of the research - There are a number of drawbacks to this study. One problem is that gender, wealth, culture, and other socio-demographic biases in selection won't be taken into consideration by either the data collection or the methods used to evaluate it. Important information on the number of patients seen by the targeted doctors was also missing from the writers' work. As a last point, healthcare administrators may be hesitant to use online data due to the difficulty in persuading clinicians to accept online remarks as actual. This study lays out a plethora of techniques that hospital administrators may use, such as sentiment analysis and time-to-next-complaint procedures, to glean valuable insights from online patient comments. This study demonstrates how administrators at medical practices and clinics may assess online patient comments using sentiment analysis and time-to-next-complaint methodologies.

Sukumar, Sreenivas Rangan et al., (2015) [56] The present movement in health IT and Big Data analytics is to create more complex models, procedures, and tools for business, operational, and clinical intelligence. But nobody is paying enough

attention to the crucial problem of data quality that these models need. With Big Data health care analytics as its backdrop, this research aims to draw attention to data quality concerns. Concept, procedure, and strategy: The analytics work done on various health data sets by different businesses is the source of the insights given in this article. Medicare and Medicaid claims, public and commercial provider enrollment databases, and electronic health records obtained via collaborations with healthcare claims processing organizations in accordance with health privacy protected standards are all part of the datasets. What we found: Here are some things to keep in mind when evaluating health care data quality: first, the full data lifecycle; second, issues caused by inaccurate or incomplete data; third, the data's origin and history; and fourth, how the original goal of data collection affects the expected results of analysis and knowledge. Technology that automates data management, storage, input, and processing has both positive and negative aspects. While automation has its uses, it also has the potential to introduce new problems with data quality. A plan that takes into account the technical and organizational components of data quality assurance is necessary for the implementation of Big Data analytics in healthcare. Consequences in practise: Data quality should primarily be determined by the value gained via analytics. With this assumption in mind, healthcare organizations that are using Big Data should be prepared to take a methodical approach to data quality. Organizations may need to develop bespoke software or data quality rule engines to address the very particular nature of health care data quality issues. novelty and worth: These days, problems with data quality are often identified and fixed piecemeal. In order to better accommodate the many facets of Big Data and the many points in the analytical process, the authors propose and lay out a data lifecycle strategy.

Dhar, Vasant. (2014). [57] There are significant barriers to broad use of predictive analytics, despite its obvious potential in health care. Using the clinical arena as an example, I go over the current status of predictive health-care analytics and how differentiating procedures might make predictive system outputs useful by promoting preventive. More individualized medical treatment is one way these systems may reduce health risks for both populations and individuals.

Greaves, Felix et al., (2013) [58] A great deal of the free-form, unstructured data on healthcare quality that is accessible online in places like social media, physician rating sites, and blogs is not collected in a systematic manner. We may be able to better comprehend and use this data to enhance healthcare quality with the help of new analytic methods like sentiment analysis. Objective Our goal was to apply machine learning to decipher patients' free-form feedback on their treatment. We classified patients' online free-text remarks on their healthcare as good or unfavorable using sentiment analysis methods. Based on the patient's free-text description and their quantitative evaluation of their treatment, through automated means, we sought to ascertain if the hospital was hygienic, whether patients were treated with respect, and whether they would personally suggest the establishment. Approach We used machine learning approaches to all 6412 online comments relating to hospitals submitted on the English National Health Service website in 2010 using the data-mining application Weka. Furthermore, we compared the sentiment analysis results with the hospital-level results of the 161 acute adult hospital trusts in England's paper-based national inpatient survey using Spearman rank correlation. Final Product For cleanliness, 84% of patients felt their remarks were addressed with dignity, and 89% said they would recommend the hospital to others based on their experience there (kappa scores: .40-.74, P).

Raposo, Mario et al., (2013) [59] Because patient happiness affects healthcare outcomes and healthcare institutions' bottom lines, it's crucial for healthcare management to regularly analyze patient satisfaction and understand what variables impact satisfaction. This study aims to examine patient satisfaction in four primary healthcare centers in Portugal by estimating a satisfaction index that will reveal which aspects of healthcare quality have the greatest impact on patient satisfaction. That was accomplished by analyzing data from 414 people to evaluate a conceptual model of primary healthcare patients' happiness. The suggested model was tested using Partial Least Squares route modeling (PLS). In a scale from 1 to 100, the findings indicate that patients' satisfaction is 60.887, which is considered to be medium. The quality of the facilities, the contact with administrative personnel, and the patient/doctor connection are the three most essential factors that contribute to patient satisfaction, in that order.

Alemi, Farrokh et al., (2012) [60] Learn how to create real-time satisfaction surveys using sentiment analysis, an AI method that categorizes sentiments expressed in text, in this article. Shorter real-time surveys are better for participation. A comment card is the quickest way to get a survey. Using sentiment analysis in this way, it is possible to generate lengthier satisfaction surveys' replies from a collection of comments. As evidence, this article presents the results of an examination of 995 comments posted on the RateMDs.com website. Doctors specializing in pediatrics and obstetrics/gynecology in the DC, MD, and VA areas were our primary emphasis. Thanks to our data, we were able to categorize patients' causes for unhappiness and find ways practices might enhance their treatment. The article details how well comments are categorized. The more feedback we get, the more precise our results will be. We also used the time-to-next-complaint metric to rank the doctors. Using a time-between control chart, we looked at whether the time-to-next-complaint was much longer than average, indicating a change from the norm. The results indicate that (1) patient feedback is readily accessible, (2) it can be categorized into positive and negative remarks using sentiment analysis, and (3) using time-to-next-complaint as a metric, we can track the effect of changes over time. It is feasible to conduct real-time satisfaction surveys according to the methods outlined in the article.

Pollettini, Juliana & Panico et al., (2012) [61] The Brazilian system of medical suggestion classifications is known as Surveillance Levels (SLs). Risk factors and patients' medical and developmental histories are used to determine SLs. Various educational and health outcomes are linked to each SL. Verifying computer-aided, automated assignment of SLs was the purpose of the current work. The software architecture of a Brazilian pediatric healthcare facility was verified and validated, and we report on it. The findings show that attribute selection significantly impacts system performance. However, by including the language module before classification, our automated suggestion of surveillance level may still take use of processing technique improvements. Our work has generalizable implications for many healthcare systems. Healthcare and government organizations may use the outcomes of systems that are based on the paper's framework to enhance healthcare services by identifying potential epidemics, developing plans to avert them, and notifying authorities to the situation.

RESEARCH GAP

Despite significant advancements in healthcare technology, there is always room for improvement in patient satisfaction and service quality. In order to assess and improve service delivery, many existing systems rely on antiquated metrics that fail to offer sufficient real-time information or to appropriately predict patients' needs. Despite showing promise in other domains, machine learning classifiers and big data analytics have received surprisingly little attention when it comes to improving the healthcare experience for patients. Recent studies have shifted their focus from patient-centered metrics like satisfaction and service quality perceptions to operational efficiency and treatment outcomes. Problems with data silos, privacy, and the lack of standardized frameworks further impede the integration of big data analytics and predictive models. To fill these gaps and improve patient outcomes, more research into the use of machine learning classifiers and big data analytics for service customization, satisfaction pattern forecasting, and healthcare delivery optimization is required.

SCSP ENSEMBLE MODEL AND FEATURE SELECTION FOR PREDICTING PATIENT HEALTH CARE OPINIONS AND HOME HEALTH CARE SERVICE

3.1 A Framework for The Analysis of Patient Opinion Systems in Health Care: SCSP

Tweets on hospital services, medical treatment, atmosphere, etc., have recently become an important part of internet reviews that people use to make judgments. Since this data is related with emotions and sentiments, its usefulness for a patient decision-making support system is crucial to its efficiency.

The enormous area of sentiment analysis is dedicated to the study of human wellbeing, which is seen as subjective as it differs according to people's needs and means of survival. It is challenging to handle the widespread anxiety in today's society over health issues, medical procedures, injuries, etc. Also, a lot of good things come out of discussions and forums on social media.

The path to responding to customer feedback by increasing goodwill may be opened by methodically examining various viewpoints. Predicting the accuracy of hidden patterns is made easier by using various machine learning and classification methods. The purpose of this study is to compare the quality of treatment at several local hospitals by analyzing their classification models, which use a machine learning strategy to predict a better model for data analysis and are used to categorize patients' feelings as positive, negative, or neutral. Based on the tweets, it seems that the healthcare business is in dire need of a trustworthy information system, which is why this particular topic was chosen.

Several Machine Learning classification models were used in the pre-processed reviews. These models were K-Nearest Neighbor, Logistic Regression, Support Vector Machine, Random Forest Classifier, and Random Forest. Using the test models, classification reports are created.

Scores for F1measure, recall, accuracy, and precision were the results. The accuracy was confirmed using Cross Validation after an examination of several models using the ensemble approach. The findings revealed that Random Classifiers exhibited the highest level of accuracy.

Using feature extraction and sentiment classifier methodologies, Section 3.1 details the methodology, data collecting, and model construction. Distinct components further subdivide the work.

The procedures for verifying the suggested method's outcomes are laid out in Section 3.2, and our results about the techniques, materials, and following discussions are presented in Section 3.3.

3.1.1 Proposed Methodology

The satisfaction of patients is of utmost importance in international healthcare. A difficulty with patient satisfaction hampered the growth of one American hospital chain, according to one study. To solve this issue, we use machine learning techniques to examine patients' opinions on the services they received.

These machine learning analyses helped them regain clients and revive their firm by identifying the problem's origin and providing quick support in rectifying the flaws.

It is more difficult than ever to meet patients' demands for additional institutions offering high-quality services due to the growth of healthcare chains and related activities. In order to ensure patient happiness, executives must address these qualities and their underlying reasons.

This research attempt aims to present some innovative methods based on the following contributions:

1. A general approach to evaluating reviews of healthcare products based on user emotion.
2. We gathered more than 1,350 Google reviews for six hospitals using outscraper.com.

3. The recommendations from the hospital are based on the relationships between the reviews.
4. Using sentiment classifier methods and ensemble approach, construct multiclass models and compare performance metrics.
5. Dividing the reviews into positive, negative, and neutral sections.

Patients may compare the hospital's value using the suggested percentage because of the reviews' polarity. Through the use of natural language processing and a wide range of criteria, the sentiment analysis challenge aims to get a comprehensive understanding of user opinion.

Various text categorization systems are used in healthcare, including positive, negative, and neutral. This is accomplished via the use of trait-based sentiment analysis techniques. The purpose of this proposed research is to analyze the relationships between written descriptions and end-user assessments. It is based on three main procedures.

- i. Cleaning and organizing the data;
- ii. Extracting and representing sentiment using a bag-of-words;
- iii. Classifying sentiment using model classifiers.

Gathering data and preparing it for analysis:

A number of medical facilities, India, have their online discussion boards mined for relevant data. Information was retrieved by means of the Outscraper.com API, which was developed using a Python script.

We moved from Excel to CSV as our rating collection tool. We have included both original and repeat reviews, but not independent reviews. We have also carefully categorized each review, including original tweets and retweets, as favorable, negative, or neutral. Six hospitals that provide a variety of specializations were approached for their opinion. The ratio of training data to test data is 80% to 20%.

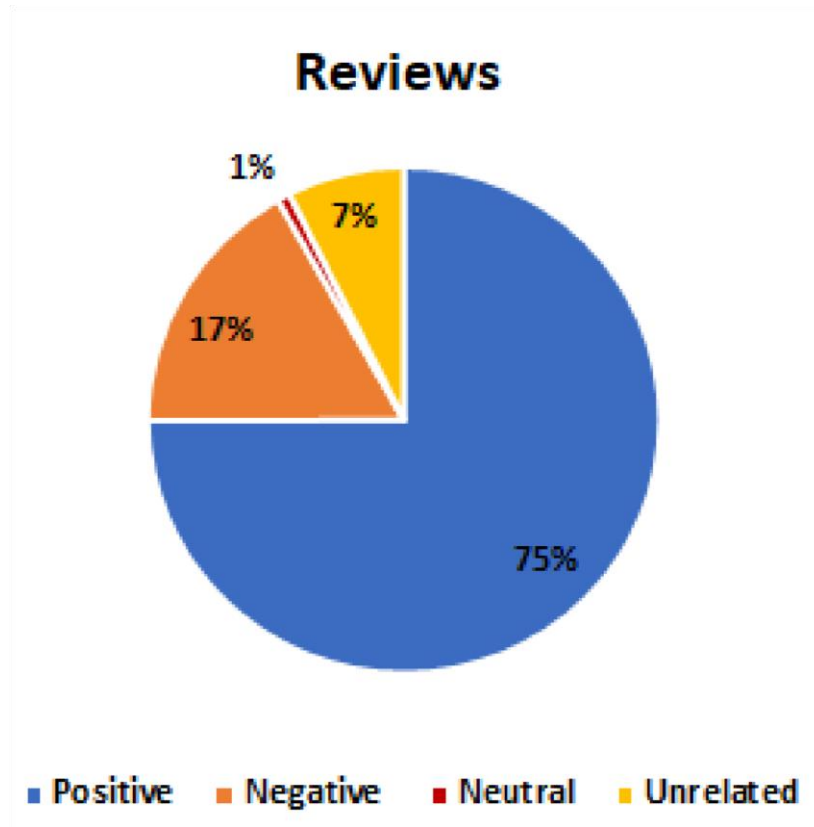


Figure 3.1: Class Distribution on Reviews

Proposed Architecture

After we delete any irrelevant reviews, we include all reviews (new and old) and carefully score each one as positive, negative, or neutral. We considered feedback from six different medical facilities.

As mentioned earlier, the data provided in Figure 3.1 is the result of preprocessing 1,350 reviews. After removing duplicates and pre-processing, the processing stage advanced to 1250 reviews, covering positive, negative, and neutral feedback.

In order to train the model and get the classification results, the classes were distributed evenly. Verifying the previous unbiased probability that caused a class distribution to be unequal was a smart move.

Out of a possible 1250 points, 1012 were deemed excellent, 227 were deemed terrible, and 11 were deemed neutral. Afterwards, we removed any assessments that

had duplicate data, URLs, punctuation, or special characters because they do not aid in emotion categorization. For a total of all reviews, see Figure 3.2.

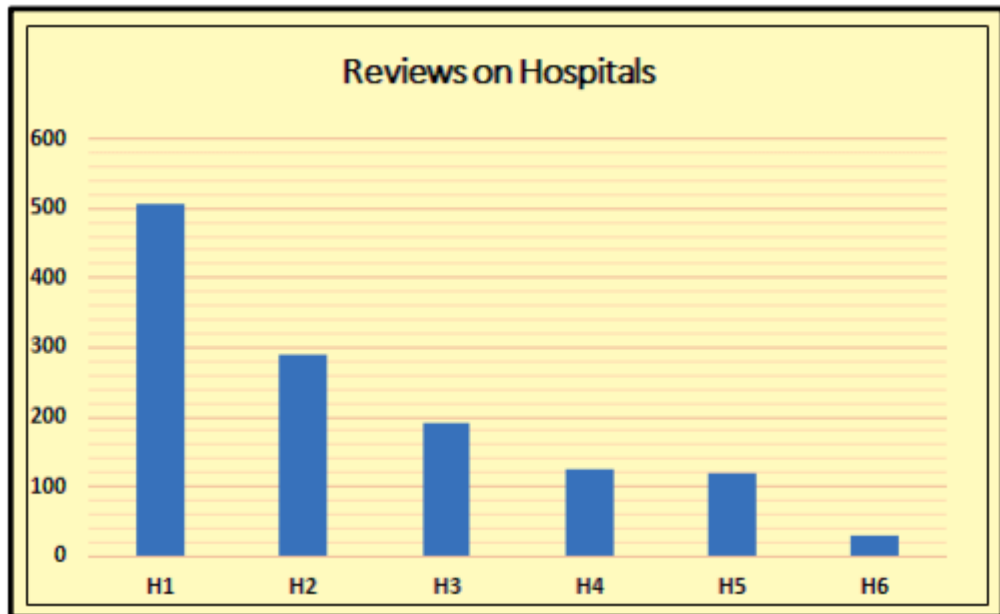
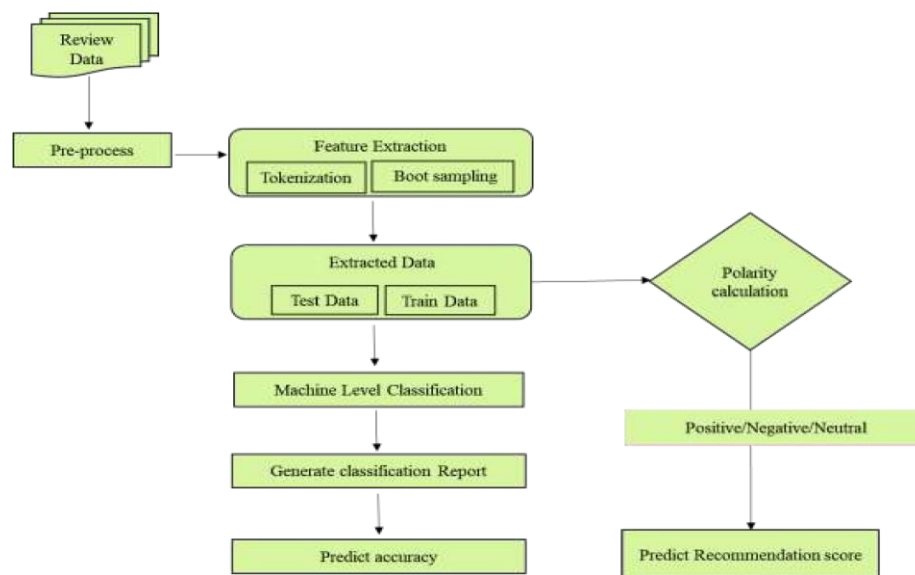


Figure 3.2 : Assessment centered on Healthcare Facilities

The below steps explain the proposed model architecture is depicted in the figure 3.3.

SCSP Model Steps

1. Begin
2. Import the data
3. Use parsing to tokenize the input data
4. Use N-gram algorithm for converting input data to Feature Vector
5. Make bag of words and samples
6. Prepare Test and Train Datasets
7. Build the Model with ML Classifiers
8. Evaluate the performance with test data\
9. Model Validation
10. Predict the Performance of ML Classifiers Precision, Recall, and F1-score.
11. Use Cross-Validation to ensure Model Accuracy.
12. End

**Figure 3.3 : Design of the SCSP Model Being Proposed**

This review lays out the suggested design in simple, pseudocode-based steps. While the first procedure demonstrates the pre-processing, procedures 2 and 3 demonstrate the feature extraction.

Procedure 1: Steps in pre-processing:

Input: Review data obtained from Review server for major hospitals

Output: Pre-processed Reviews

Parameters:

R_g : Regular Expression in python library, R_w : Review, G_w : Get word, T_g : Tag,

T_{gd} : Tag Data, P_{rw} : Parse Review

$T_g = R_g$

$T_{gd} = \{$

$R_g.1[Adjectives]$

$R_g.2[Noun]$

$R_g.3[Verb]$

$R_g.4[Adverbs]$

$\}$

$P_{rw} = T_{gd}.get [T_g[0]]$

If P_{rw}

return P_{rw}

else

return 3

To construct a word corpus, tokenizing the data is done after pre-processing using feature extraction. As could be anticipated, social evaluation material does not make use of formal English, meaning it does not adhere to proper grammar, punctuation, etc. In order to resolve this issue, the stemming approach was used to build relationships between the several words that were repeated. Both "likes" and "liked" have their origins in the English word "like." It helps reduce feature sparsity as well.

Unigrams and bigrams are features that text emotion classifiers may employ. As an alternative to bigrams and unigrams, N-grams are being considered since they may provide more information about the general mood. The N-gram, the Corpus Approach, stop words, and lemmatization are some of the ways that weighted features vectors might be represented. Reviews benefit from the N-gram method as it employs a set window size of 'n' words to concentrate on.

Two code samples are provided by the review tokenization/lemmatizing pseudo code process and three code examples are provided by the bag words pseudo code process.

Procedure 2: Steps in Tokenization using lemmatizer

Input: *Parsed Review data obtained from Pre-process Review server for major hospitals*
Output: *Tokenized Reviews*
Parameters:
R_g: Regular Expression in python library, R_w: Review, P_{rw}: Parse Review, C_R column Review P_s: Porter stemmer, w = word
Initialization: *P_{rw} <-[NULL]*
Begin *R_w in C_R do:*
 X = load[R_w]
 P_{rw} <- X[R_w]
End
For each *R_w in P_{rw} do*
 R_w = R_g.sub('[a-z,A-Z] ' for each C_{RW}[R_w[i]]
 R_w = R_g.lowercase[R_w]
 R_w = R_g.split[R_w]
End
Create main root stem for each word in R_w
 R_w =[P_s.stem[w] for each R_w [w] if not w in set(stopwords.w('English'))
 R_w = ' ', join R_w
End

It is possible to extract characteristics from reviews by following the pre-processing processes described in procedure 2. We begin with the preprocessed data and, after following these steps, we get the tokenized data that was derived from it.

Procedure 3: Creating Bag of words using corpus

Input: *Tokenized Review data obtained from Tokenization*
Output: *cleaned Review word ready for model*
Parameters:
C_R: Corpus,, R_w: Review
Initialization:
 C_R <-[NULL]
For *I in range(0, R_w)*
 If(i+1)%100==0):
 C_R.append [R_w]

As a third stage, we used Python's Bag-of-words module in conjunction with stop words extraction to remove articles. Here we will go over the second and third phases of the approach outlined before in greater detail.

1. One must first deconstruct text into its individual words and spaces before proceeding with tokenization. This method is known as bag-of-words. Monosyllabic terms such as "can't," "that's," and "he'd" were, however, ensured to be there.
2. "a," "an," and "the" are articles that should not be removed from the Bag-of-Words.
3. Third, using a string of consecutive words, construct an N-gram collection. As negations, the words "no" and "not" are appended to the words that come before or after them. It is possible to construct the following bigrams from the sentence "I do not like service": "I do not", "do not like", and "not like service". Given the significance of negation in articulating emotion and viewpoint, this approach allows for more exact classification.
4. The angles of vectors generated by the counting vectorization and lemmatization functions in Python have been calculated using the cosine similarity. This method may be used to find the similarity of check phrases that represent vectors. We then find the sentiment polarity using the hospital's name and reviews, and then we suggest that specific hospital based on that.

Before and after procedures 2 and 3, the lemmatization process using stop words is shown in Figures 3.4 and 3.5, respectively.

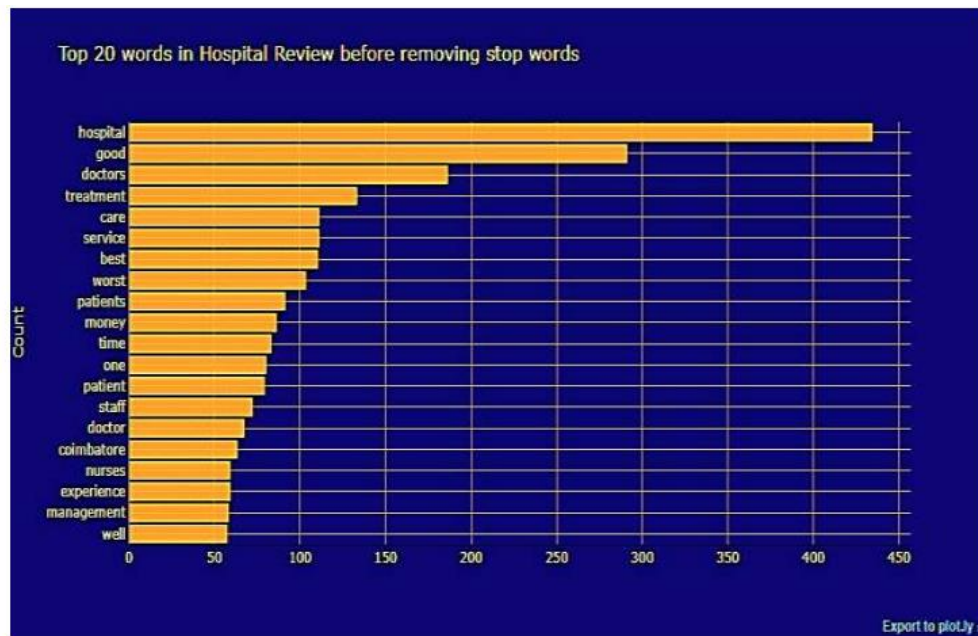


Figure 3.4: Top 20 terms used in hospital assessments prior to the elimination of stop words

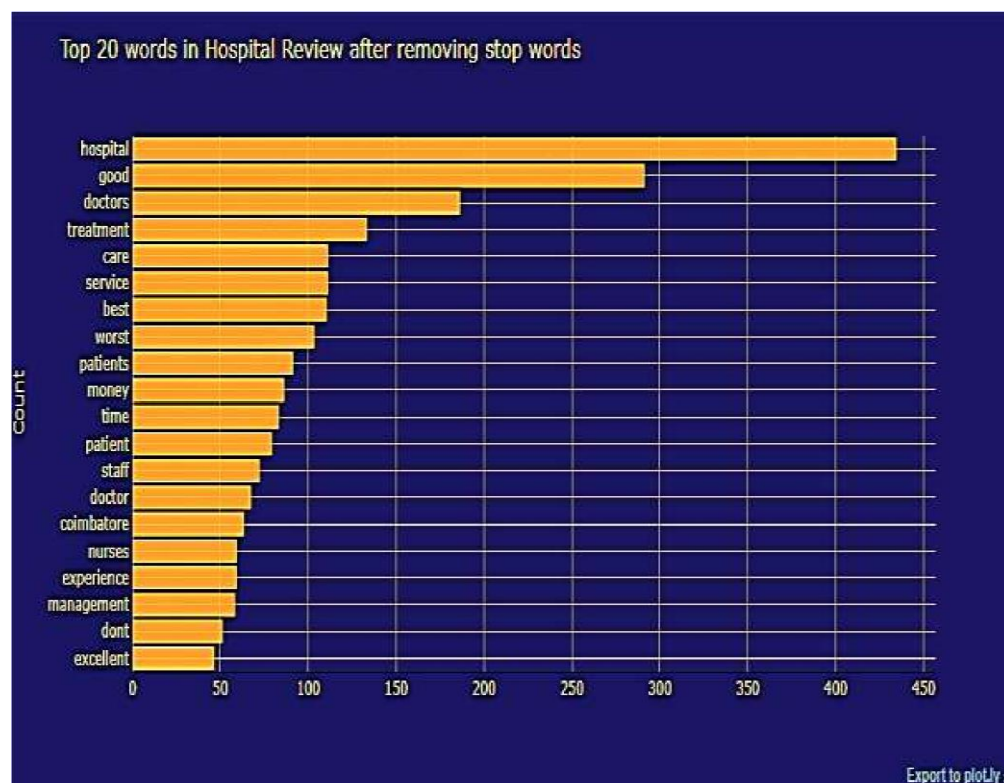


Figure 3.5 : The top twenty terms removed from the hospital review

Figure 3.6 and Figure 3.7 demonstrate two instances of bigram usage before stop words and after stop words, respectively. Bigrams combine two words into a single one.

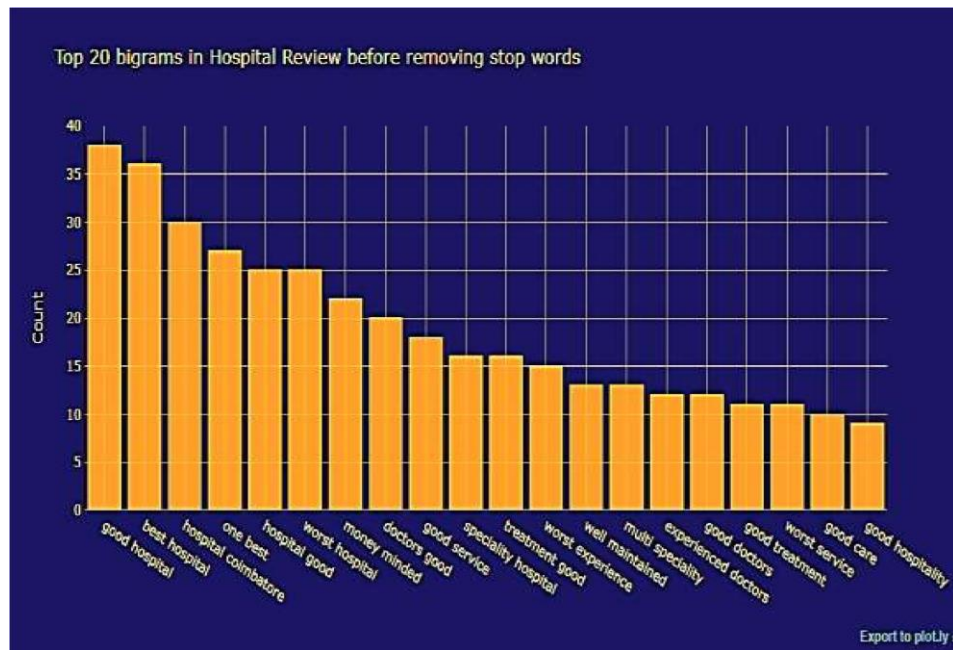


Figure 3.6: Prior to removing stop words, the top 20-bigrams in hospital reviews

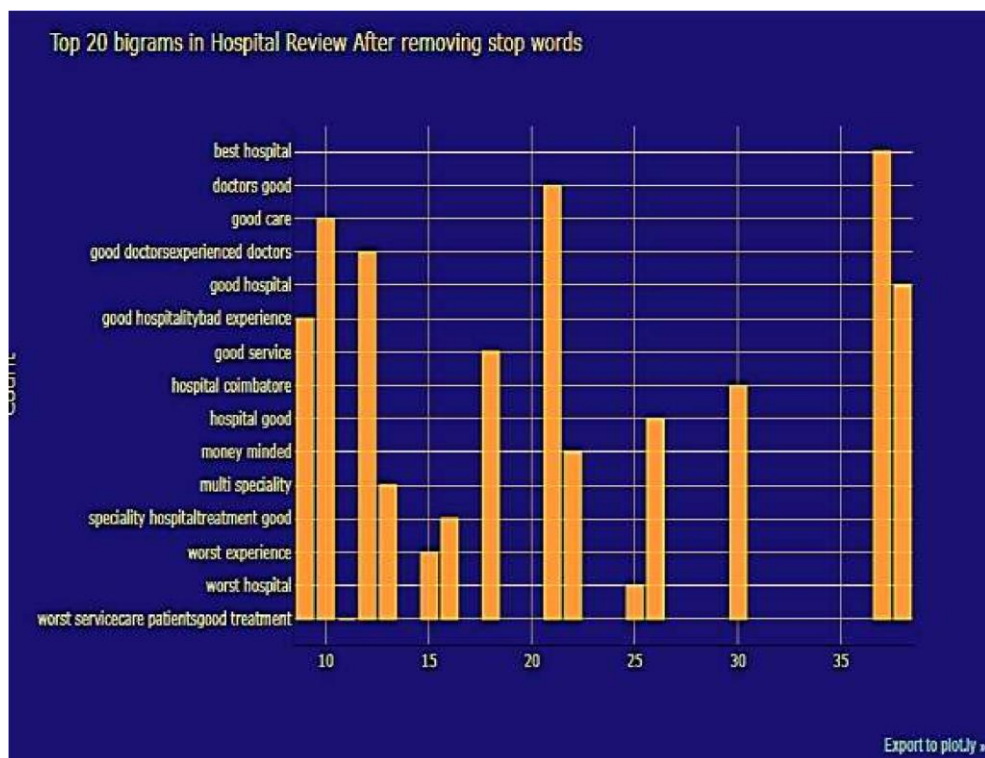


Figure 3.7: Twenty most important terms removed from hospital evaluations

Figures 3.4–3.7 show, graphically, the representation of text data during feature extraction in procedures 2 and 3.

SCSP Classification Approaches:

Once the pre-process and functions have been collected, machine learning classifiers are used for sentiment analysis. In this study, we compared and contrasted other studies and employed the following methods based on our findings. DT, RF, NB, KNN, LR, SVM, and RF

- **SVM**

Using a cutoff level to differentiate between data points in a binary classification issue, support vector machines (SVMs) are a common tool for emotion analysis. We will only take into account the threshold that differentiates the most valuable margin classes, regardless of whether many thresholds present. We used this model into our investigation due to its strong performance in sentiment analysis.

- **LR**

When the result is always a 0 or 1, LR may be used to solve classification problems.

In our investigation, we set the threshold using a 0.5 probability.

- **NB**

A lot of people think highly of the NB model for sentiment analysis. An application of the Bayes rule is shown by the following:

We have $P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$ as the probability of X given Y.

Using this idea, we constructed a word vector using frequencies and put it through its paces.

- **KNN**

The KNN classification method is extensively used in sentiment analysis as a pattern identification tool. We may normalize the attribute values from the Euclidean distance by assigning greater weight to qualities with a narrower range. Its 86% success rate in prior research has made it a popular machine learning classification strategy for sentiment analysis.

- **RF**

A classification tree with binary outcomes and a probability estimate for every class member is the end product of the RF method. It employs a method known as bagging, which involves reusing the original sample rather than the feature sample. For the model's fine-tuning, the Scikit-learn toolkit's maximum tree depth was set to 200 and 400 estimators, respectively. We put the XGBoost decision tree model through its paces here. We won't consider it, however, since the outcome will be the same.

- **Multi-classification –SCSP Ensemble Model**

In the research, seven different categorization models were used. We have used cross validation and 10-fold validation to evaluate SVM, LR, NB, RF, DT, and KNN, which are machine learning classifiers. Table 3.2 shows the results of the three-class classification experiment, whereas Table 3.1 shows the results of the five-class classification experiment.

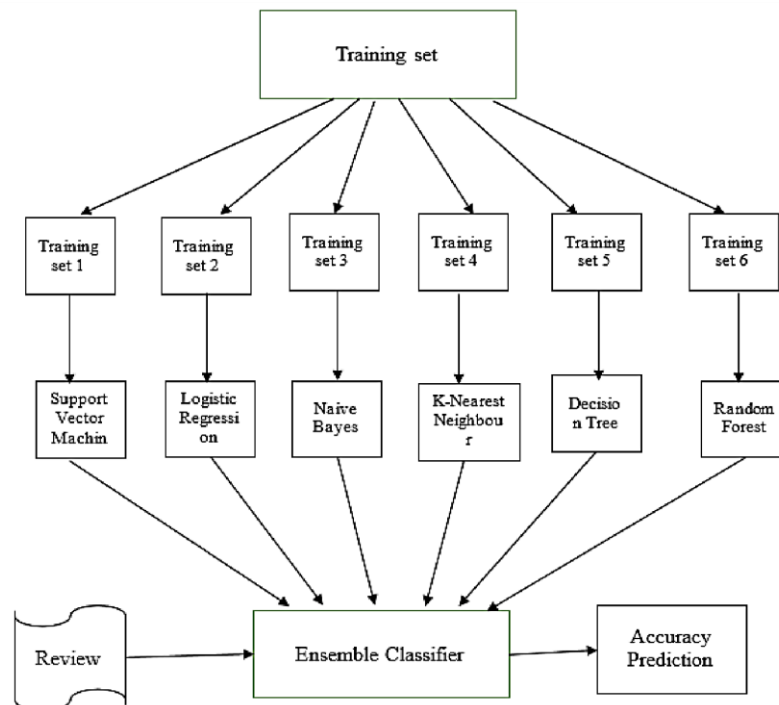
Two of the weakest classifiers, DT and bagging, were similarly accurate, therefore we didn't include them in our analysis. When used with the six sentiment classifiers, the ensemble model can forecast the training dataset's overall performance. Figure 3.8 below displays the outcome of training the different parts separately.

**Table 3.1: Comparative Study of Machine Learning Classifier Accuracy on a
Five-Class Dataset**

Accuracy: 5-class dataset						
ML Classification Model	SVM	LR	NB	RF	KNN	DT
Classifiers - Accuracy	63	64	65	60	59	60

Table 3.2: Accuracy of ML Classifiers in a Three-Class Set

Accuracy: 3-class dataset						
ML Classification Model	SVM	LR	NB	RF	KNN	DT
Classifiers - Accuracy	90	90	87	96	93	65

**Figure 3.8: An ensemble model for SCSP**

Using the majority vote approach for Bootstrap sampling, the SCSP Ensemble model classifies the reviews (Figure 3.8). Currently, we are taking into account all

six classifiers that have the same weights. The outcome predicts the outcome for each class according to the number of votes it gets from all six classifiers, even if these classifiers rank scores sequentially. In Section 4, the pseudocode for the majority vote is shown.

Procedure 4: Creating voting using Majority voting method

Input: Processed Review data obtained after processing

Output: Predicted accuracy for each model

Parameters:

CR_w : classification results for Review, S_R : Sentiment Classes Result, $CR_{wi} = class1$, $CR_{wj} = class2$ $CR_{wk} = class3$

Initialization:

For i in S_R and CR_{wi} , CR_{wj} , CR_{wk} in CR_w

$CR_{wi} > 3$

then $class = CR_{wi}$

Else if $CR_{wj} = 3$

then $class = CR_{wj}$

Else if $CR_{wk} < 3$

then

$class = CR_{wk}$

return $class$

Procedure 4 is used to classify the evaluated data when none of the six potential classification outcomes can be anticipated using the majority vote approach. Put yourself in the shoes of a reviewer and picture this: SVM rates it positively, LR negative, DT neutral, KNN positive, and NB negative. Plus, RF and NB both think it's great. After then, one of the three groups might be given the review. We trained each of the six classification models independently using 10-fold cross validation on the identical dataset. Python 3.6 is used for both the model's training and classification processes. Cosine similarity allows us to classify reviews as positive, negative, or neutral. We may then provide an informed recommendation on the hospital based on this.

3.1.2 Discussion and Findings

Figures 3.4 and 3.7 show that the reviews are first turned into pre-processed data using N-grams and stop-words. A positive, negative, or neutral classification of the dataset could emerge from the sentiment analysis. By using cross-validation, we were able to sidestep the common problem of over-fitting that arises when training

with a limited data set. Here, you may use one set of data to train the model, and then use the other set to see whether it was accurate. Our data subsets were split into ten equal halves in order to accommodate the 10-fold validation. After training the model with nine data subsets, we used the residual subset to assess its performance. When all of these validation steps are combined, they provide an average. We contrast two sets of data, one with five and the other with three.

There are five separate quality levels in the five-category dataset: excellent, good, neutral, bad, and dreadful. The three-category data set only allows for positive, neutral, and negative categorization. The 3class is the result of merging the 5class's good and terrible groupings. The 3-class data set's accuracy scores are shown in Table, while the 5-class data set's values are shown in Table.

Recall, precision, and the F-measure were among the measures used to assess the model's accuracy. The top model was obtained by cross-validating the accuracy measures of many machine learning models. The results of this study are not particularly convincing when contrasted with previous conclusions in sentiment analysis. All sentiment classifiers that used cross-validation before ensemble and class binding procedures have their classification reports shown in Table 3.3 and Figure 3.9.

Table 3.3: Prior to ensemble, model classification

	Accuracy of SVM: 0.6370967741935484		
Sentiments	Precisions	Recall	F1-Score
Positive	0.83	0.83	0.83
Negative	0.45	0.47	0.46
Neutral	0.60	0.57	0.59
Accuracy of Logistic Reg.: 0.64516129032258 06			
Positive	0.83	0.85	0.84
Negative	0.45	0.42	0.43
Neutral	0.59	0.62	0.60

Accuracy of Naïve Bayes: 0.87142857142857 14			
Positive	0.91	0.95	0.93
Negative	0.84	0.77	0.80
Neutral	0.85	0.88	0.86
Accuracy of KNN:0.5967741935483871			
Positive	0.70	0.83	0.76
Negative	0.47	0.53	0.50
Neutral	0.57	0.40	0.47
Accuracy of Random Forest: 0.6048387096774194			
Positive	0.81	0.83	0.82
Negative	0.36	0.25	0.30
Neutral	0.54	0.67	0.60

The ensemble technique yields better review performance accuracy.

The dataset is supplemented using a random sampling method. You can see from Table 3.4 that we guarantee a high degree of correctness. See Figure 3.4 for the pre- and post-ensemble findings.

Table 3.4: Model classification after Ensemble

	Accuracy of SVM: 0.9047619047619048		
Sentiments	Precisions	Recall	F1-Score
Positive	0.99	0.96	0.97
Negative	0.80	0.90	0.85
Neutral	0.92	0.85	0.88
Accuracy of LR: 0.90			
Positive	0.99	0.96	0.97

Negative	0.79	0.89	0.84
Neutral	0.91	0.85	0.88
Accuracy of Naïve Bayes: 0.87142857142857 14			
Positive	0.91	0.95	0.93
Negative	0.84	0.77	0.80
Neutral	0.85	0.88	0.86
Accuracy of KNN:0.7619047619047619			
Positive	0.82	0.91	0.86
Negative	0.64	0.77	0.70
Neutral	0.83	0.60	0.69
Accuracy of Random Forest: 0.9285714285714286			
Positive	1.00	0.97	0.99
Negative	0.88	0.92	0.90
Neutral	0.90	0.89	0.90

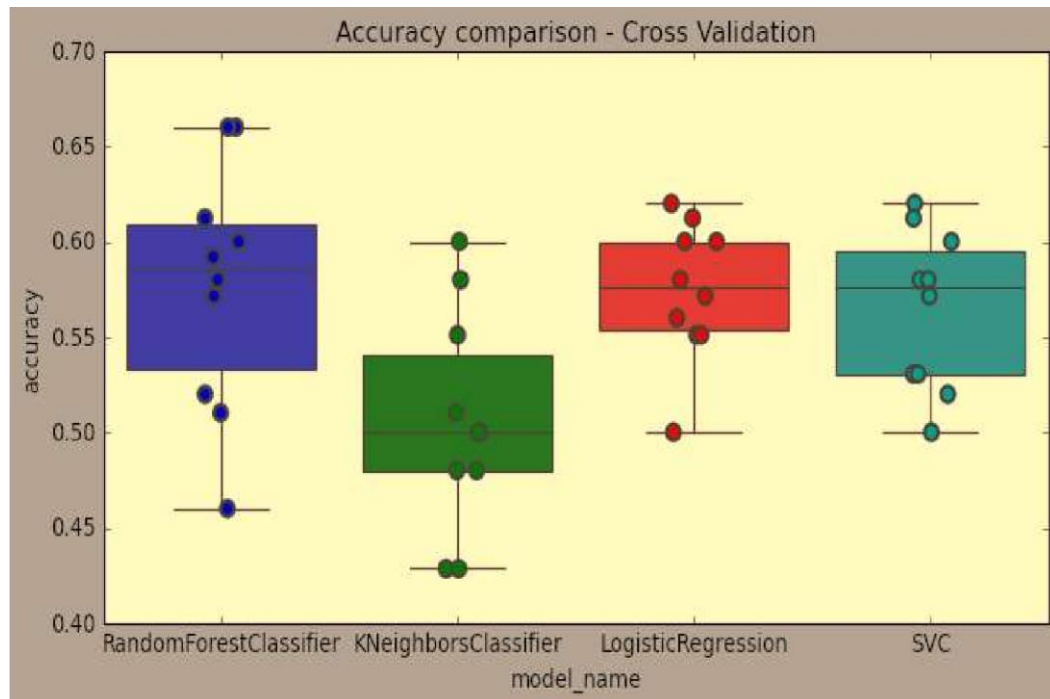


Figure 3.9: Comparison of Model before ensemble

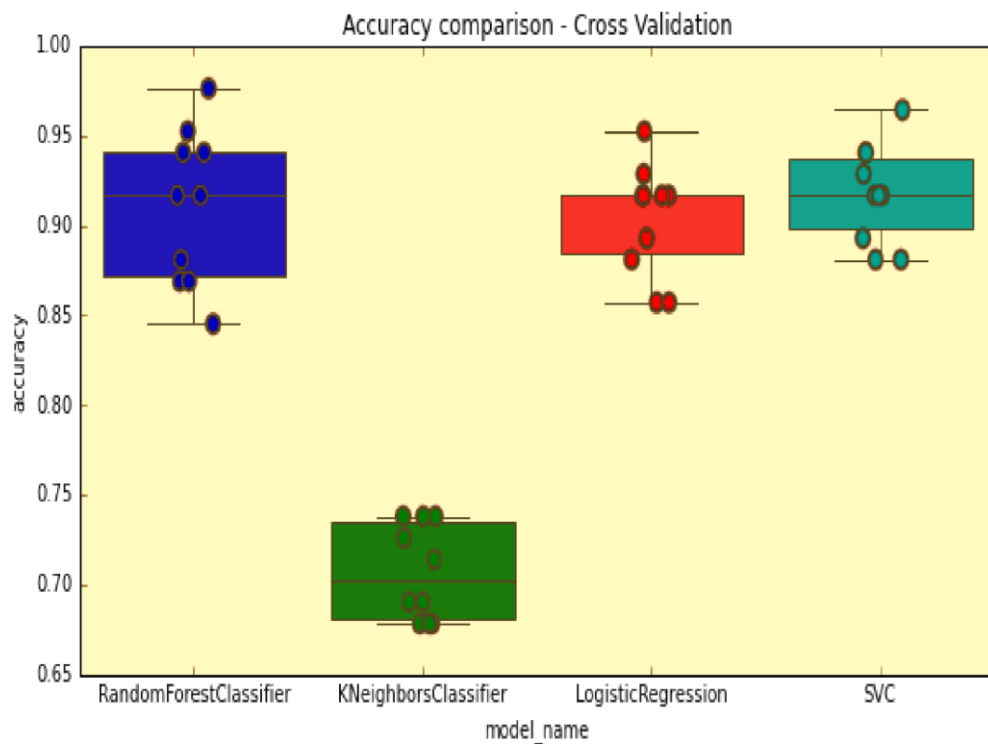


Figure 3.10: Comparison of Model after ensemble

Tables 3.3 and 3.4 show that out of all the classifiers, Random Forest has the highest accuracy at 93%. Consequently, this is the best way to gauge how the general population feels about health care.

3.1.3 Model Validation

A confusion matrix allows for four potential outcomes: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). In the table below, you can see the results of the models' classification performance as evaluated by several ML approaches. To find out if sentiment classification was feasible, we utilized F-measure metrics for accuracy, recall, and cross-validation to evaluate many models. Tabulated in Table 3.5 are the results of the ensemble method mood classifier model for each classifier model. The results are confirmed when the ROC curve of several models' predictions is examined.

Placing the classifier toward the upper left corner enhances the model's performance. Using a variety of color codes, Figure 3.12 displays many emotion classification systems. With a ROC-AUC of 95% and an accuracy of 93%, our findings demonstrate that the RF model outperforms its competitors.

Table 3.5: Classifier Model scores based on reviews

Classifier Model	F1	Accuracy	Precision	Recall	AUC_ROC
SVM	97	90	99	96	93
LR	97	90	99	96	85
NB	93	87	91	95	88
KNN	86	76	82	91	81
RF	99	93	100	97	95

ROC CURVE

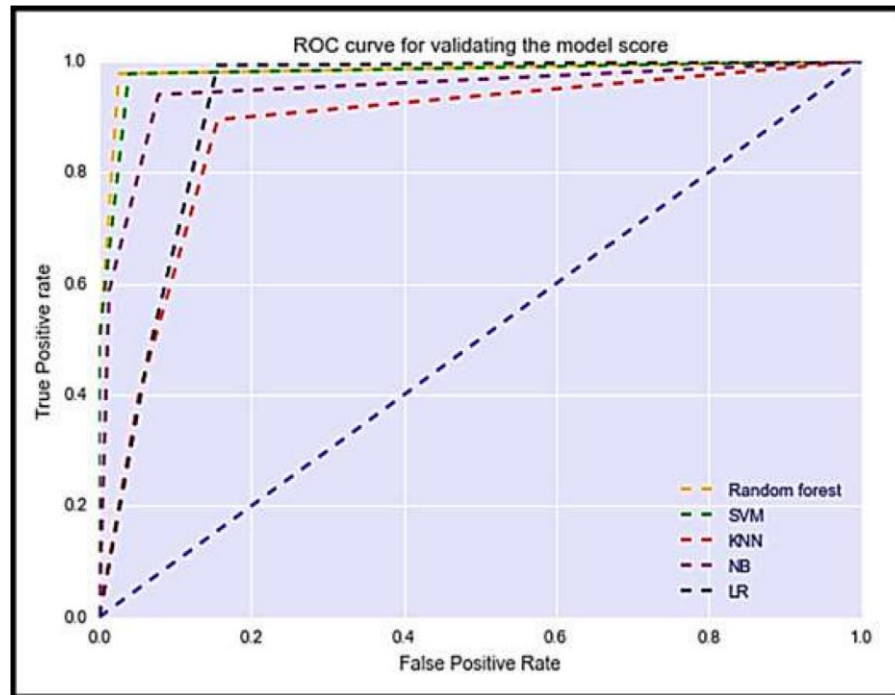


Figure 3.11: ROC curve

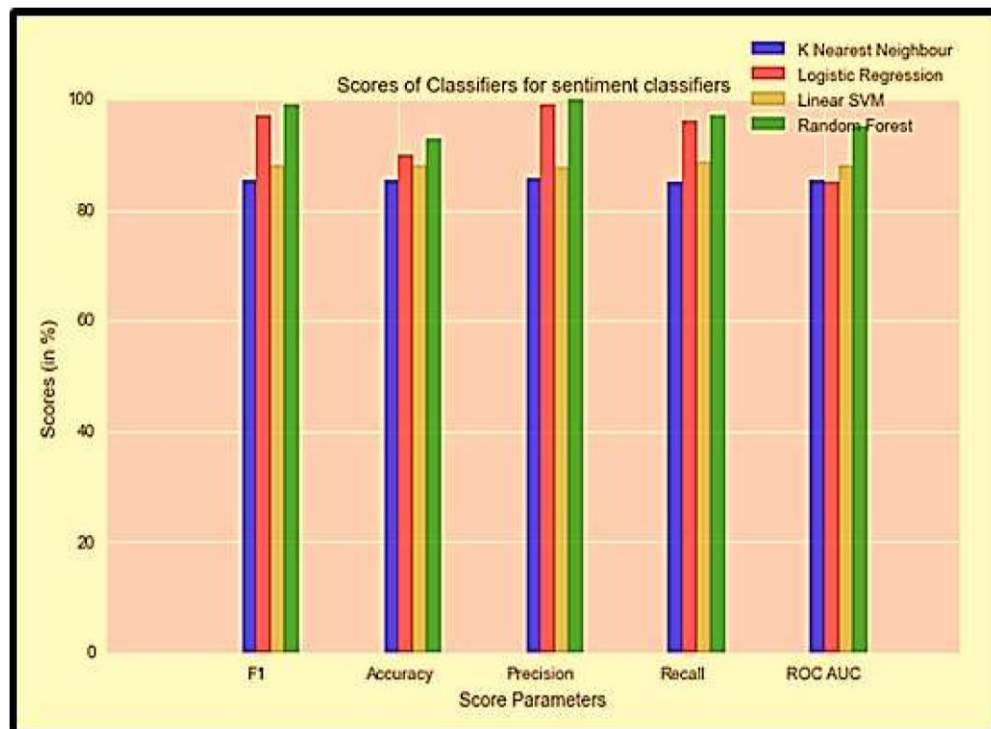


Figure 3.12: Prediction Metric Score

3.1.4 Novelty and Summary

At various threshold levels, the AUC-ROC accuracy is used to assess the function and classification model. Furthermore, it provides the methods by which each model may assess separability. A bigger area under the curve (AUC) indicates better model prediction. Figure 3.12 and Table 3.5 show the prediction metrics for the emotion classifier model.

The results show that Support Vector Machines (SVM) outperform all other Random Forest Sentiment Classifier models. With a 95% AUC and a 93% SVM accuracy, it surpasses previous studies. The proposed new model is built upon this basis; it will assess healthcare-related social media reviews with an AUC accuracy of 95% and an overall accuracy of 93%.

Table 3.6 shows the predicted percentage of outcomes that are recommended after a good, negative, or neutral evaluation. The unique model will choose a credible hospital according to the review's polarity. Presented below are some advic.

Total number of positive reviews/total number of reviews * 100

Table 3.6: Hospital recommendation

S.No.	Hospital Name	Positive	Negative	Neutral	Recommendation %
Class 1 Hospitals with the highest percentage					
1	H1	424	80	1	83.96039604
2	H5	98	18	2	83.05084746
3	H3	155	35	1	81.15183246
4	H4	102	21	3	80.95238095
Class 2 Hospitals with the second highest percentage					
5	H6	23	3	3	79.31034483
6	H2	219	70	1	75.51724138

3.2 Feature Selection for Predicting Home Health Care Services

Keep an eye on treatment quality if you're an investment in the healthcare business. There are a lot of ways that healthcare organizations gauge their patients' happiness. The health of the people is one of the top priorities for every government body. There are several aspects to consider when assessing a service's quality, including its efficacy, safety, communication, and general care. Opinion mining is greatly used for assessing the quality of healthcare. It seeks for patterns that patients may have concealed in order to measure consumer contentment. The burden of determining the quality of healthcare services is firmly placed on the patient in this case. In recent times, customer opinions on all things pertaining to healthcare have been evaluated via patient evaluations on internet forums or response forms provided by the hospital facility. This allows companies to expand their customer base and enhance their products and services. Since they describe real people's experiences, reviews posted by real patients are the best way to judge healthcare quality. Previous studies found a significant correlation between patient satisfaction and healthcare business growth. Moreover, prior studies have shown that identifying the most critical component of patient satisfaction is an enormous undertaking.

Much research remains into the factors associated with opinion mining and how they affect patients' levels of satisfaction. Thus, the question of which feature metric best measures patient happiness remains open. Multiple dimensions allow for the prediction of a multitude of quantitative traits, which in turn allows researchers to reach a multitude of definitive conclusions. According to the research, there aren't enough models and tools to measure this degree of contentment. The most common kind of assessment is a survey that asks patients to rank their overall satisfaction with a service. It would be fantastic if researchers dug more into the question of how to quantify patient pleasure in a way that makes doctors stand out. Finding the many opinion mining aspects that are associated with various metrics for measuring patient satisfaction using conventional methods is difficult. 'In order to assess patient happiness, machine learning classification techniques like LR, RF, GB, and AB are used to find possible features. These characteristics aid in determining the quality of the portion. A crucial step in evaluating healthcare service quality and patient satisfaction is the identification of an additional metric for measuring treatment

quality according to patient perspectives. The Healthcare Assessment and Quality of Life Survey (HCAHPS) is used to assess patients' experiences in the US.

It generates reliable findings for hospital evaluations and provides several insights for improving service quality. In order to evaluate and contrast the home health care services offered by various suppliers, HCAHPS carried out a survey in several cities throughout Australia. This research proposes an ML-based method for evaluating patient rating datasets. Through the use of feature importance, the most crucial aspect for patient happiness may be identified. Model categories are documented in a report. The ROC curve is a useful tool for determining if a model is accurate while doing cross-validation. Furthermore, the following sections provide more detail on the methodology and data collection that will be utilized. They also outline the steps to validate and analyze the model's performance. Lastly, they draw conclusions from this model.

3.2.1 Proposed Methodology

Here, we take a look at the suggested approach

Framework

The framework as a whole recognizes that the purpose of the feature selection approach is to choose the most useful attributes for making model predictions. In order to classify home health care data, the proposed research makes an informed assumption based on the selected features. It also advocates for a novel method of classification that makes use of ML models—specifically, a feature importance model and a feature selection model. Figure 3.13 displays the proposed framework. Reading the data into Python, then filling in missing values and applying a filter to format it are all part of data preparation. Using the feature importance, we can obtain robust prediction using the assessed model. Afterwards, we use an ML classifier to choose the optimal approach by virtue of the distinct characteristics we retrieved.

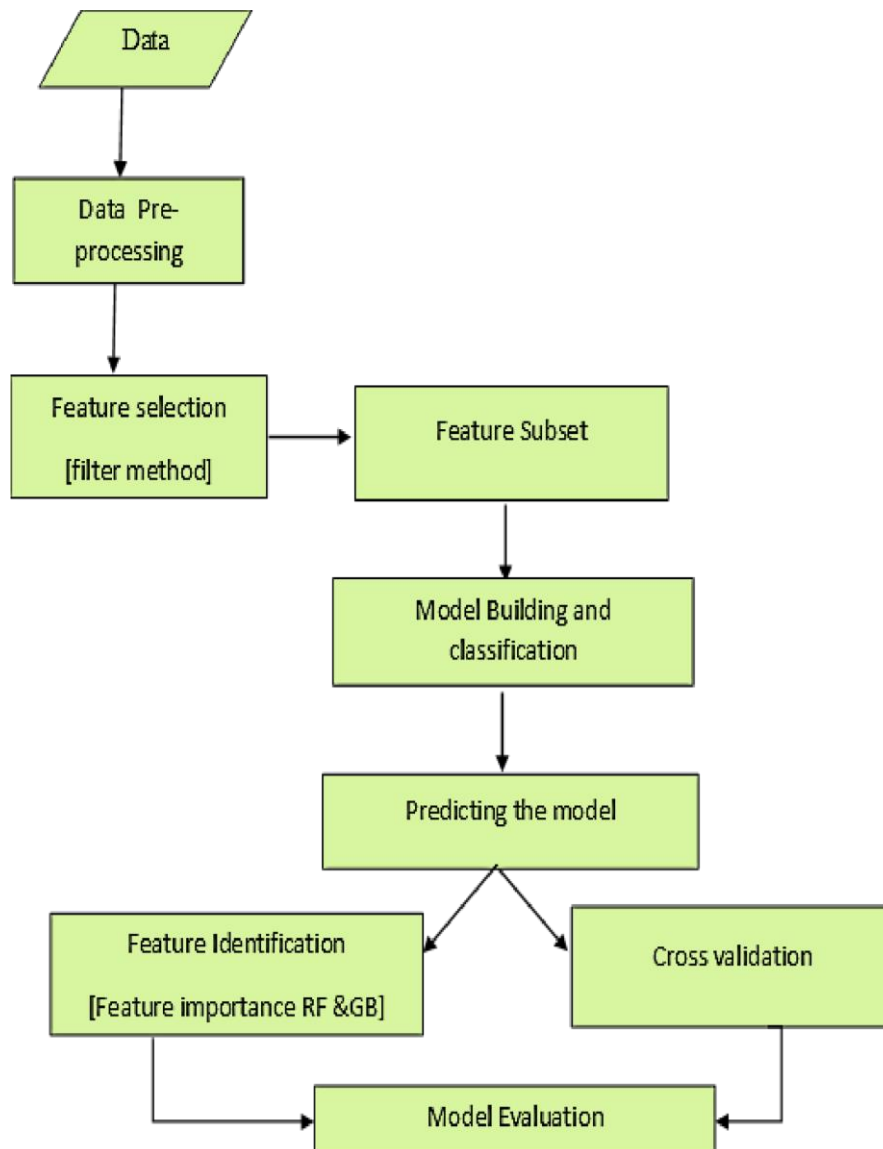


Figure 3.13.: Proposed Framework

Implementation

What follows is a rundown of the methodology and the proposed framework for home health care prediction using a filter-based feature selection approach, building on the previously stated strategy.

The sequence of procedural steps

1. Begin
2. Import data for home healthcare.
3. Implement Data Pre-Processing.
4. Create the Feature Subset FS.
5. Gather test and training data.
6. Build the model RF, GB and AdaBoost
7. Generate Feature Importance using RF, GB and AdaBoost
8. Using the ML classifiers RF, GB, and AdaBoost, train and build models with a feature subset using repeated K-fold cv. • Produce a variable importance score
9. Train and build models using the feature subsets generated by Feature Importance with repeated K-fold cv and the ML classifiers RF, GB, and AdaBoost.
10. Using test data, evaluate the performance of all models.
11. Predict the best feature subset and classification strategy.
12. End

Data

The necessary information is taken straight from the HCAHPS website. To determine which home health care services are most effective, it compiles electronic health records from several Australian locations.

HHS, MSS, SP, OT, PT, and NC are just a few of the many services offered by these organizations. A number of prescribed measures are eliciting responses from patients.

This research aims to examine, using the provided metrics, which one patients consider most important when deciding how much to pay for medical treatment. Only organizations that provide full services were considered for inclusion in the study, and a total of 12,165 answers were obtained.

We did not include the agency in our research since it does not provide the services listed above. Eliminating these constraints allowed for the examination of 11,365 replies.

Table 3.7: Total number of respondents

Total Response	Offers all the services	Agency that does not offer all the services
12165	11365	800

Data Pre-Processing

The chosen data set is substituted with the medians of the columns. The median is the most successful imputation strategy when dealing with outliers and missing values since it is the middlemost value. There are a total of 11,365 occurrences in the data record after preprocessing. The data is further separated into a training set and a test set to make model creation and performance evaluation easier.

Feature Selection

As feature selection decreases the dataset's content, it is essential for obtaining the best model prediction with adequate accuracy. By using this approach, computers may discover the most relevant features in a dataset without any human intervention. The feature data is used to determine the rank in ranking-based feature selection. When data size increases dramatically, it's impossible to avoid increasing the computing budget. Overfitting, which reduces the model's performance, could occur if the data set contains inappropriate characteristics. When building economic models, feature selection is a useful approach for reducing the number of superfluous attributes. The cleansed dataset seemed to have 12,165 rows and 14 columns at first glance. Data cleaning reduced the number of rows from 14,165 to 11,365 columns, as seen in Table 3.8. We reduced the total number of rows to 11,365, but maintained missing values greater than the threshold and deleted them to match the research's inclusion criteria. A description of the data and metrics used to assess the function

can be found in Table 3.10. To find the most valuable feature, we utilized a mixture of univariant and bivariate analysis using a heap map, and a variance threshold of 0.7. Then, we used the filter strategy. In order to get the intended outcome, feature selection is used to obtain model classification and evaluation.

Table 3.8: Feature description before selecting the features

	CMS							
	Certification Number (CCN)*	Zip	Phone	Target	Professional	Communication	Safety	Overall care
			1.215900e+04			12159.000000		12159.000000
count	12159.000000	12159.000000		12159.000000	12159.000000		12159.000000	
mean	326171.615182	57901.661732	6.028284e+09	3.843573	3.948680	3.990377	3.035858	3.071305
			2.498248e+09	0.673474				
std	226214.421878	25335.435766			0.697915	0.687938	0.735017	0.762739
min	17000.000000	601.000000	2.012916e+09	1.000000	1.000000	1.000000	1.000000	1.000000

25 %	117094.0 0 0000	35963 .50 0000	3.308 363 e+09	4.000 000	4.00000 0	4.000000	3.000 000	3.000 000
50 %	267655.0 0 0000	60607 .00 0000	6.182 779 e+09	4.000 000	4.00000 0	4.000000	3.000 000	3.000 000
			8.181 149 e+09	4.000 000				
75 %	459232.5 0 0000	77478 .00 0000			4.00000 0	4.000000	3.000 000	3.000 000
max	747974.0 0 0000	99901 .00 0000	9.898 949 e+09	5.000 000	5.00000 0	5.000000	5.000 000	5.000 000

Table 3.9: Feature description after exclusion of features

	Professional	Communication	Safety	Overall care	Target
count	11365.0000 0	11365.00000	11365.0000 0	11365.0000 0	11365.0000 0
Mean	3.917466	3.995952	3.048042	4.001496	3.934800
Std	0.773925	0.700343	0.761152	0.728256	0.679221
Min	1.000000	1.000000	1.000000	1.000000	1.000000

25%	4.000000	4.000000	3.000000	4.000000	4.000000
50%	4.000000	4.000000	3.000000	4.000000	4.000000
75%	4.000000	4.000000	3.000000	4.000000	4.000000
Max	5.000000	5.000000	5.000000	5.000000	5.000000

Table 3.10 Features based on sentiments

	Professional			Communication			Safety			Overall Care		
	Pos itiv e	Ne utr al	Neg ativ e	Pos itiv e	Ne utr al	Neg ativ e	Pos itiv e	Nea utra l	Neg ativ e	Pos itiv e	Nea utra l	Neg ativ e
Comb inatio n gover nment volunt ary	19	2	0	20	0	1	5	12	4	20	0	1
Hospi tal based progr am	706	37	6	725	16	8	378	334	37	711	31	7
local	746 3	74 8	558	772 0	52 5	524	131 6	640 3	105 0	759 9	672	498

Offici al health agenc y	102 4	12 0	76	106 5	73	82	195	850	175	103 9	114	67
Skille d nursin g facilit y based progr am	32	7	7	33	7	6	6	27	13	33	6	7
Visiti ng nurse associ ation	444	81	35	487	44	29	140	342	78	474	55	31

3.2.2 Model Building and Classification

Applying ML classification algorithms to datasets in order to forecast accuracy scores based on features chosen allows for the categorization of the best HHCS from among the available agencies.

The dataset is trained and validated using popular classification algorithms such as RF, GB, and AdaBoost. In the section under "Results and Discussion," the findings are thoroughly examined.

Feature Importance

It is a method for predicting the target value by giving a score to the input information. Aid in reducing input characteristics while providing detailed insight into the data and model. Model predictions made by an ML classification system inform the weighting of features.

Cross validation

Applying several ML classification algorithms to the dataset helps identify the most effective attributes for HHCS categorization. When training on a limited data set, over-fitting might occur; to avoid this, K-fold cross-validation is used to train the model and find the best model using several classifiers. One set of data is used to train the model, while the other set is used to test the model. This is the method behind the separation of sets. We divided the data subsets into ten equal halves in order to do 10-fold validation. Nine data subsets were used for training the model, while the remaining one was used for validation. This validation technique typically yields ten models. The models are built using three ML classification techniques—RF, GB, and AdaBoost—along with the given attributes. One may find a detailed report on the analysis and discussion of the model's output in the results section.

Performance Evaluation:

To validate the models' performance, which is evaluated using an ML classification approach, we look at how accurate the outputs are. When testing algorithms like RF, GB, and AdaBoost, it is necessary to use training data with features chosen using a filter method. Next, the selected model's accuracy is used to predict the most valuable features, and the resulting model is evaluated for feature weights and relevance based on those features. The predicted accuracy model is built from these data and evaluated using ROC curve.

3.2.3 Findings and Discussion

In this section, we elaborate lavishly on the findings that follow from our suggested technique. Data preprocessing is carried out after data purification in order to determine the optimal function taking into account the various aspects. This section

presents the results of our study that were derived using the technique that was suggested.

Feature Subset

In order to improve our classification method, feature selection involves finding the most useful features from the available ones and removing the unnecessary or redundant ones. Classification accuracy and correct prediction are both negatively affected by characteristics that lack important information. Finding the optimal feature is possible using a variety of approaches. To put this into practice, we used techniques like the correlation matrix and the elimination of columns with a threshold higher than 0.7.

Table 3.11: Feature score for scatter and correlation matrix

	Profes- sional	Communica- tion	Safety	Over- all care	Recom- mend	Target
Professional	1.000000	0.778357	0.546364	0.803867	0.765208	0.879129
Communica- tion	0.778357	1.000000	0.553225	0.748145	0.725218	0.854973
Safety	0.546364	0.553225	1.000000	0.553629	0.564403	0.728596
Overall care	0.803867	0.748145	0.553629	1.000000	0.785103	0.864424
Recommend	0.765208	0.725218	0.564403	0.785103	1.000000	0.789398
Target	0.879129	0.854973	0.728596	0.864424	0.789398	1.000000

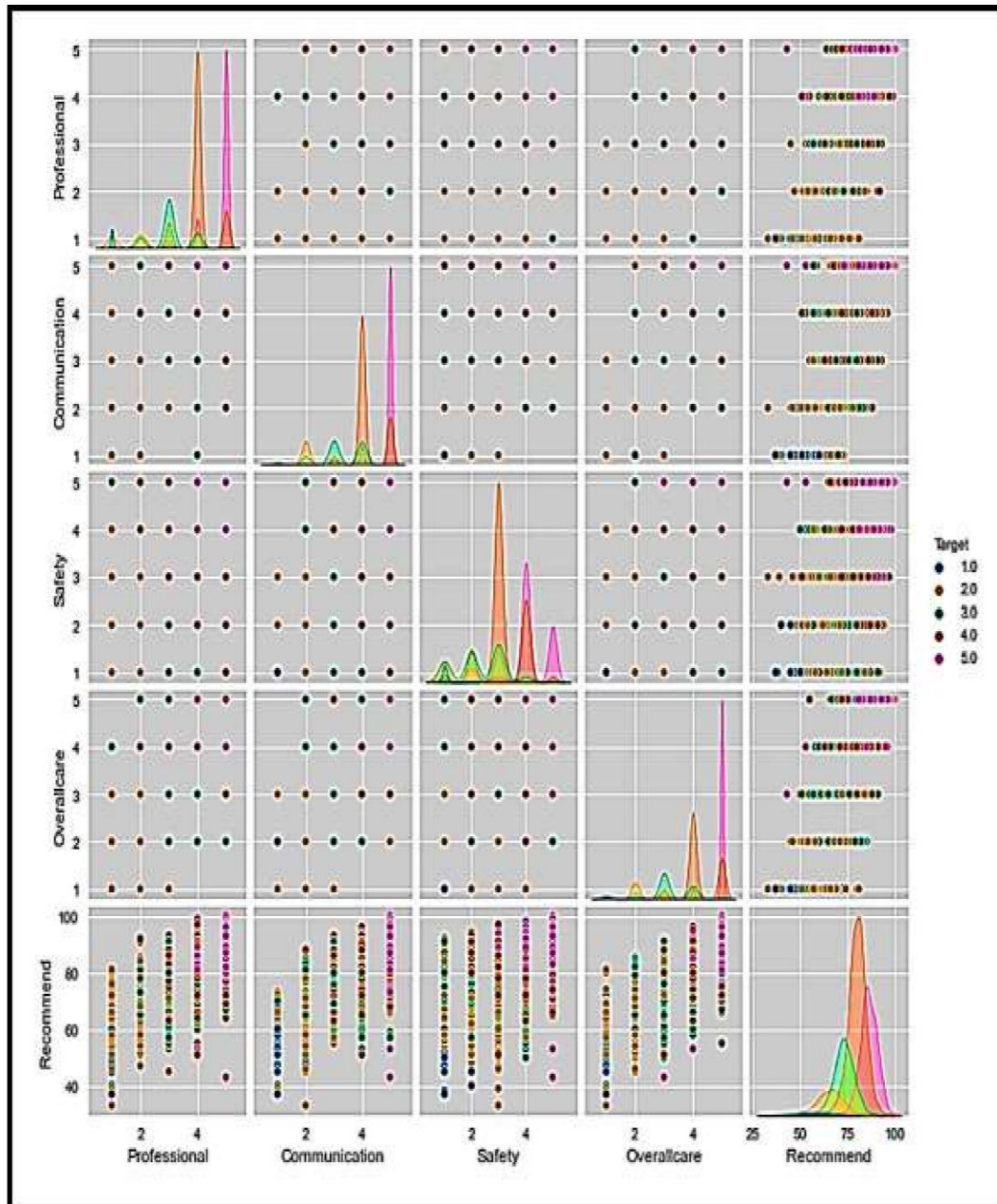


Figure 3.14: Scatter Plot

Figure 3.14, which is a scatter plot, provides further information about the data. In some instances, the characteristics reveal a positively skewed distribution, while in others, it is more direct.

This suggests that these traits are connected to outcomes and may be dependent on one another.

To have a better understanding of the characteristics, we may examine the correlation matrix and heat map.

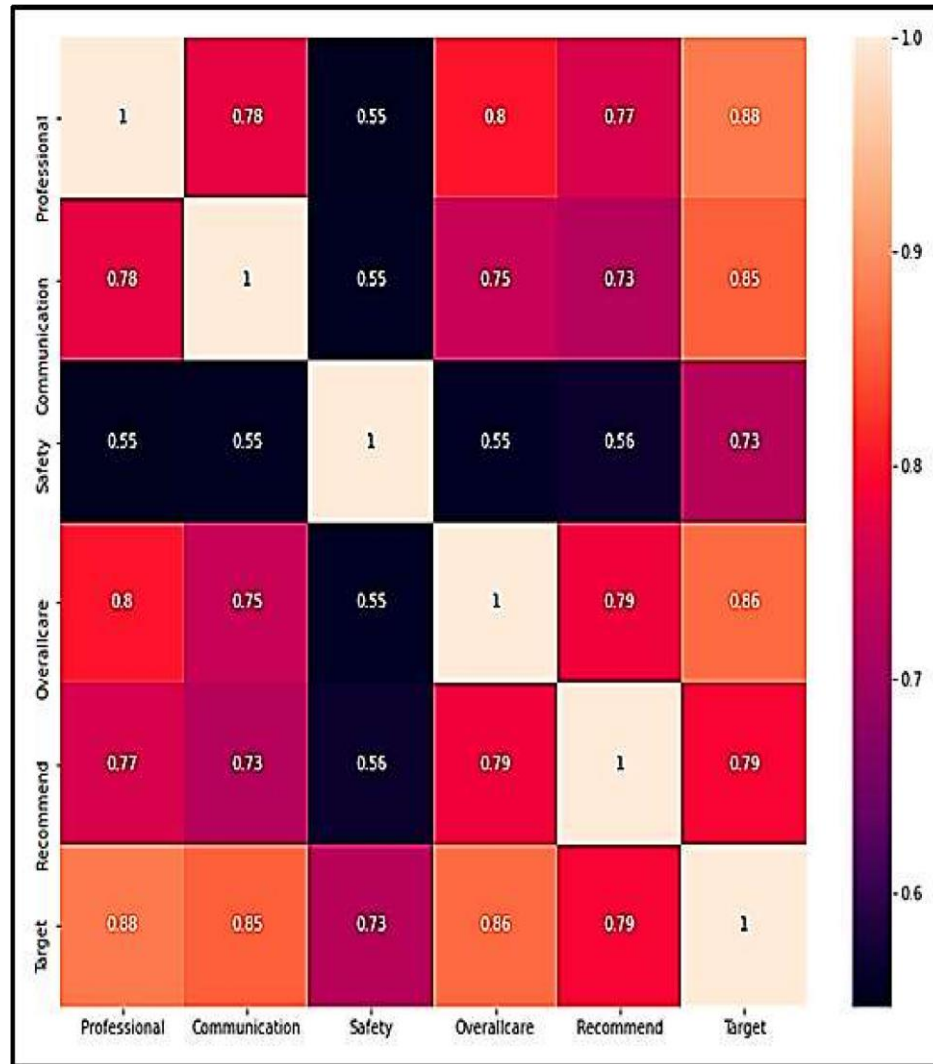


Figure 3.15. Correlation Matrix

Table 3.11 and figure 3.15 demonstrate that there is a clear link between the characteristics and the outcomes for squares with positive values. The ratio depicted in orange becomes more considerable as the value increases. When one function goes up, the other one goes up as well; when one goes down, the other one goes down as well. An inverse correlation is shown by the negative squares. In conclusion, there is evidence that these qualities are dependent on one another.

Patients rated the services they received according to a number of factors, including communication, overall care, professionalism, and safety, as shown in

Table 3.12 and Figure 3.16. As previously mentioned, HHCS provides a wide range of services according to patients' ratings on a 5-point scale (with 5 being excellent, 4 being very good, 3 being decent, 2 being awful, and 1 being very bad).

Table 3.12: Ranking based on features

S.No.	Professional	Communica- tion	Safety	Overall care
1	290	65	604	98
2	392	585	753	513
3	995	665	2237	878
4	2246	2335	1573	1930
5	1711	1984	467	2215

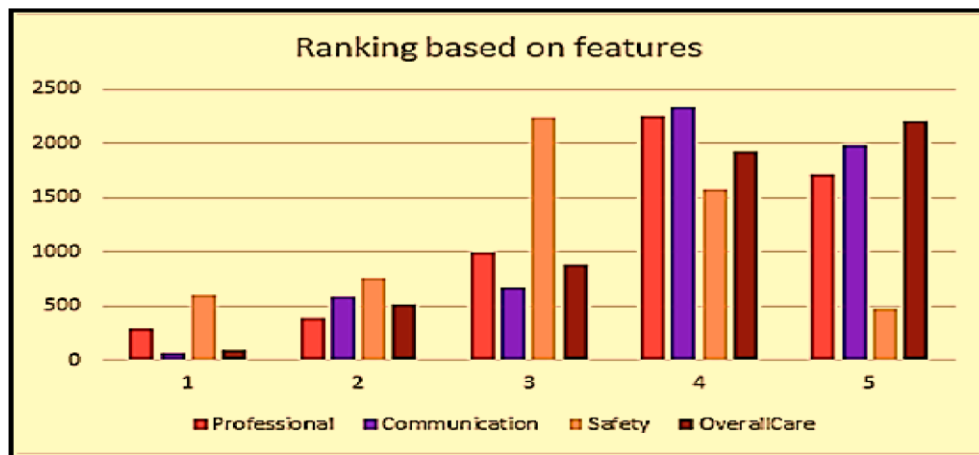


Figure 3.16: Ranking based on features

Table 3.12 and Figure 3.16 show that most services are rated 4 or 5, depending on which feature is desired. The majority of ratings for the chosen qualities are 4, whereas the majority are 3, for safety. The most highly rated aspects are professional, communication, and overall care.

3.2.4 Machine Learning Classifier

Logistic Regression

Logistic When solving a binary classification issue using regression, the resulting models can only take on two values: zero or one. Model performance recall was 98% and accuracy was 98.01 using the chosen feature to forecast patient happiness with regard to various HHCS. Figure 3.17 and Table 3.13 display the results.

Table 3.13. Logistic regression classification report

Logistic Regression Accuracy : 98.01563493181			
	Precision	Recall	F1-Score
1	0.98	0.98	0.98
2	0.98	0.97	0.98
3	0.97	0.98	0.97
4	0.98	0.98	0.98
5	0.98	0.98	0.98



Figure 3.17: Classification report of Logistic regression

Random Forest

The decision tree, a hierarchical structure based on features, is the foundation of the RF classifier. Each node in the decision tree is categorized based on a subset of its related qualities. This set of decision trees was constructed using samples from several datasets. Here, we classify the nodes using the Gini index of the feature subset.

Table 3.14 Random Forest

Random forest Accuracy: 99.86801583809944			
	Precision	Recall	F1-Score
1	1.00	1.00	1.00
2	1.00	0.99	1.00
3	0.99	1.00	0.99
4	1.00	1.00	1.00
5	1.00	1.00	1.00

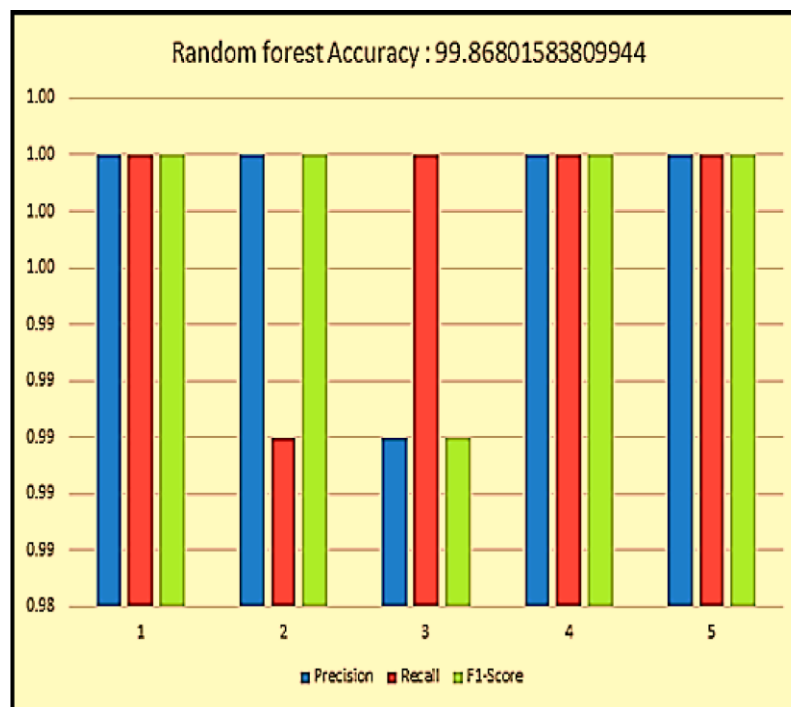


Figure 3.18: Classification report of Random Forest

Gradient Boost

Training several underperforming learners to enhance each other's performance is the goal of this machine learning classifier method. A new tree is constructed at each stage by correcting the errors of previous trees. Obtaining this result was made possible by the best machine learning approach for healthcare applications.

Table 3.15: Gradient boost

Gradient Boost Accuracy: 99.8240211174659			
	Precision	Recall	F1-Score
1	1.00	1.00	1.00
2	1.00	0.99	1.00
3	0.99	1.00	0.99
4	1.00	1.00	1.00
5	1.00	1.00	1.00

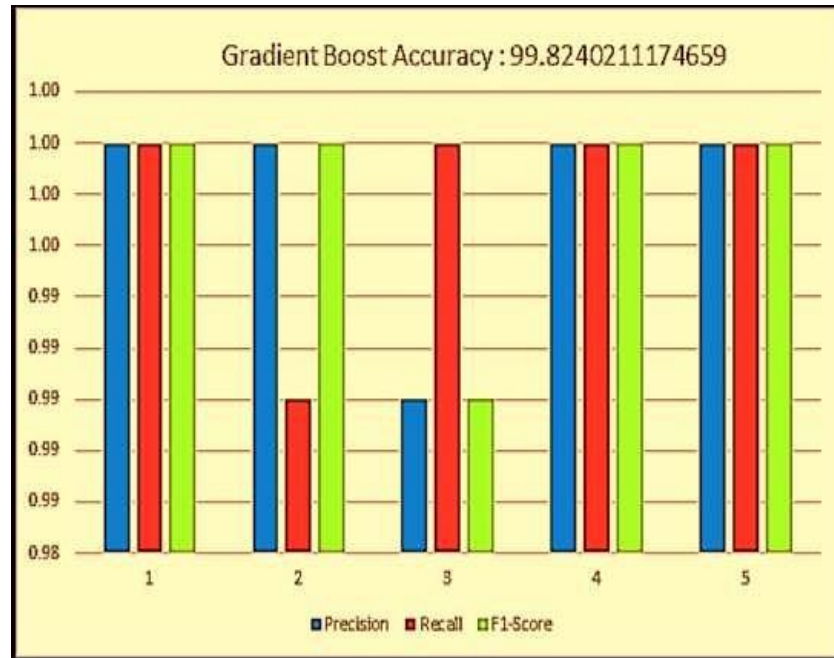


Figure 3.19: Classification report of Gradient Boost

Ada Boost

It is well-known that AdaBoost is a machine learning classifier. Utilizing the majority voting method and the weak classifier together, the boosting approach is used. To a large extent, the final classifier's voting scheme's weights determine the training accuracy. Among ML classifiers, one that makes use of weights from recent studies outperforms the majority.

Table 3.16. Ada Boost

Ada Boost Accuracy: 86.8675758908931			
	Precision	Recall	F1-Score
1	1.00	1.00	1.00
2	1.00	0.99	1.00
3	0.99	1.00	0.99
4	1.00	1.00	1.00
5	1.00	1.00	1.00

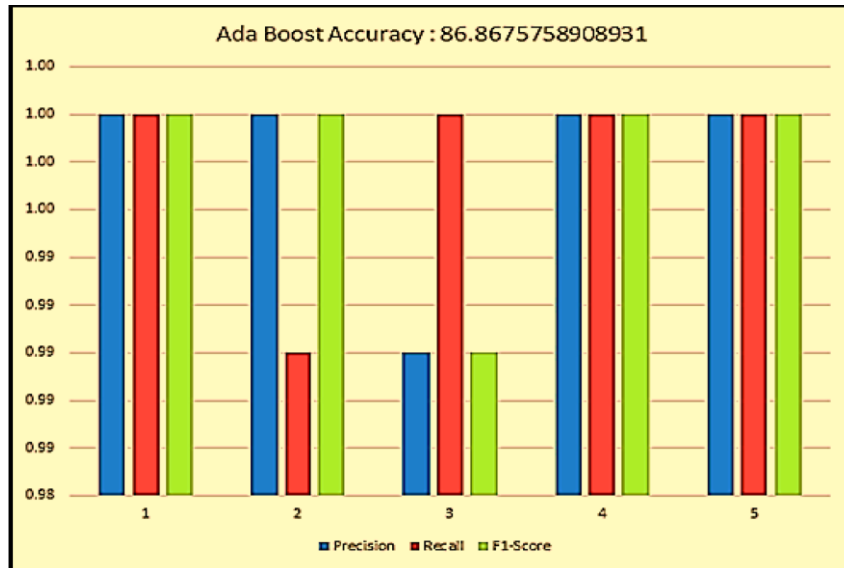


Figure 3.20: Classification report of Ada Boost

3.2.5 Feature Importance

Feature variables are shown on the y-axis, while feature scores are shown on the x-axis. Professionalism, Communication, Safety, and Overall Care were determined to be key aspects in patient satisfaction based on the score. Among these factors, HHCS patients were most satisfied with the professional service they got, followed closely by the safety they felt they were well protected.

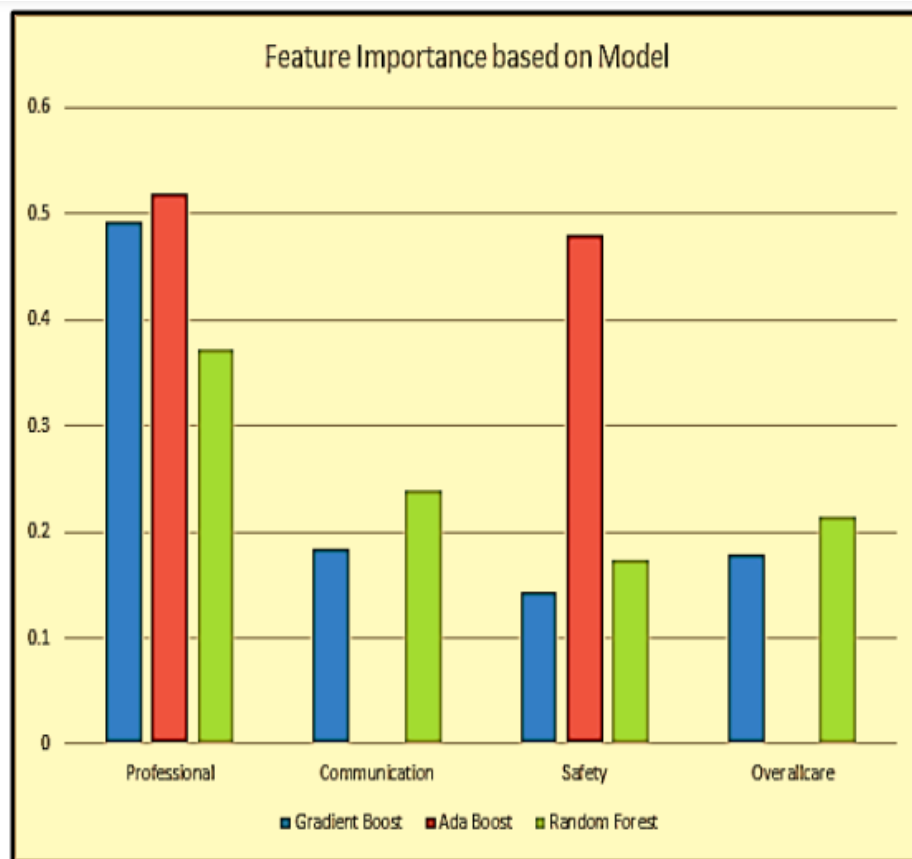


Figure 3.21: Feature Importance based on model

3.2.6. Model Validation

K-fold Cross validation

Cross validation is a method for training and validating ML models that compares several models. We use the accuracy score to train and validate models such as LR, RF, GB, and ADB in our work. To avoid over-fitting, which may occur when training with a small data set, the model uses cross-validation. This method divides the data into two parts: the training set and the validation set. Since we used 10-fold validation, we divided the data subsets into ten equal halves.

The results of the cross-validation accuracy comparison are shown in Figure 3.22.

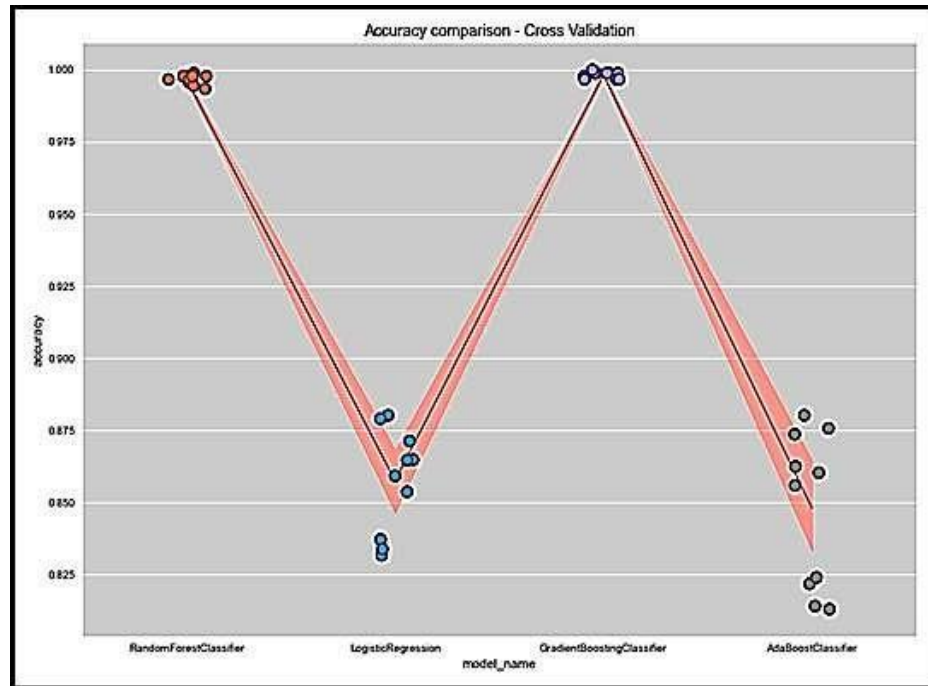


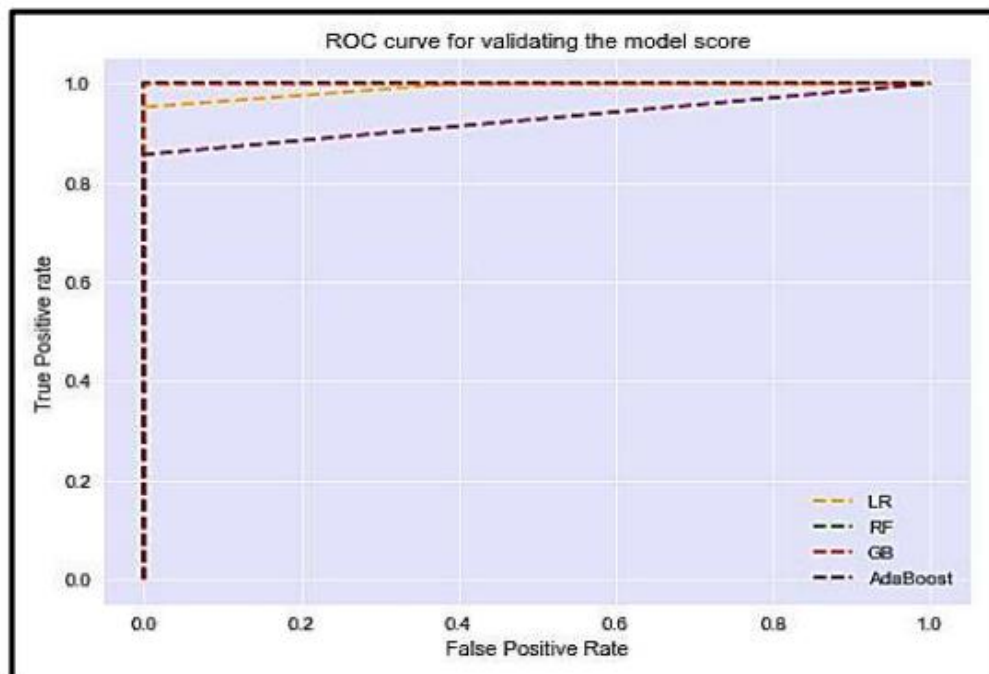
Figure 3.22: Model Comparison

Model Performance Measure

Predicting the model's performance and other performance measures are used to assess machine learning classification algorithms. Making use of many ML classification algorithms—including LR, RF, GB, and ADB—allows us to predict the model's performance. The performance indicators included in this study are evaluated by computing the ROC curve value. Table 3.17 shows that among the HHCS models, RF and GB are the most patient-friendly. Model performance as measured by ROC values is shown in Figure 3.23.

Table 3.17: Model performance using ROC

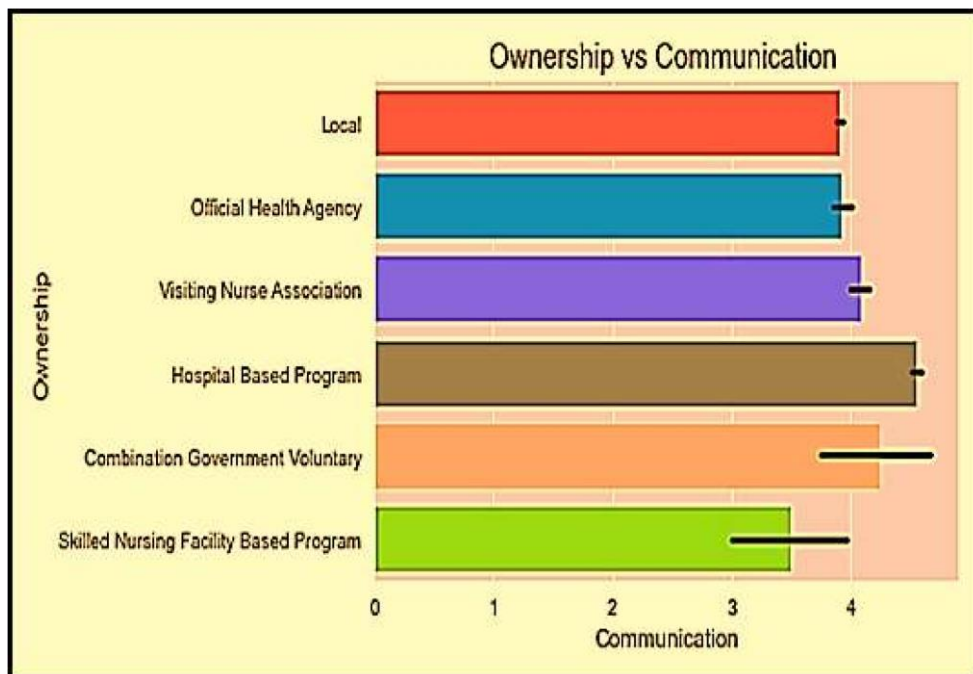
Models	ROC
LR	0.758933127315834
RF	0.9960827310675624
GB	0.9972869097127411
AdaB	0.733487225430504

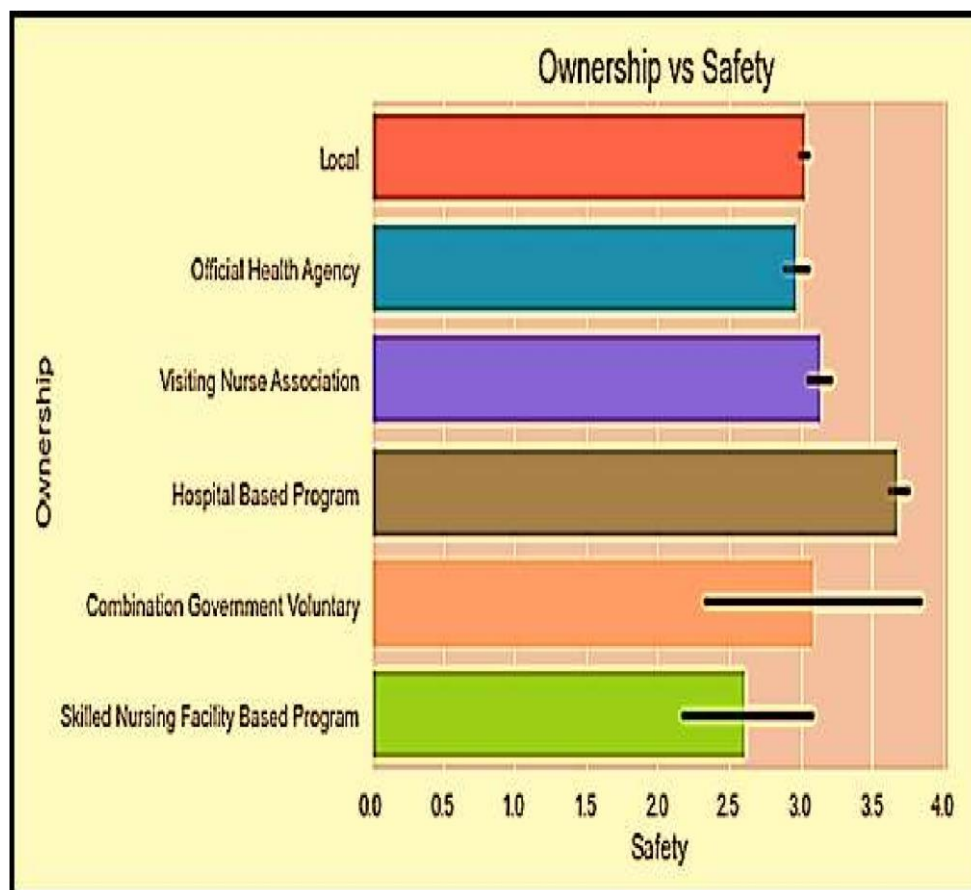
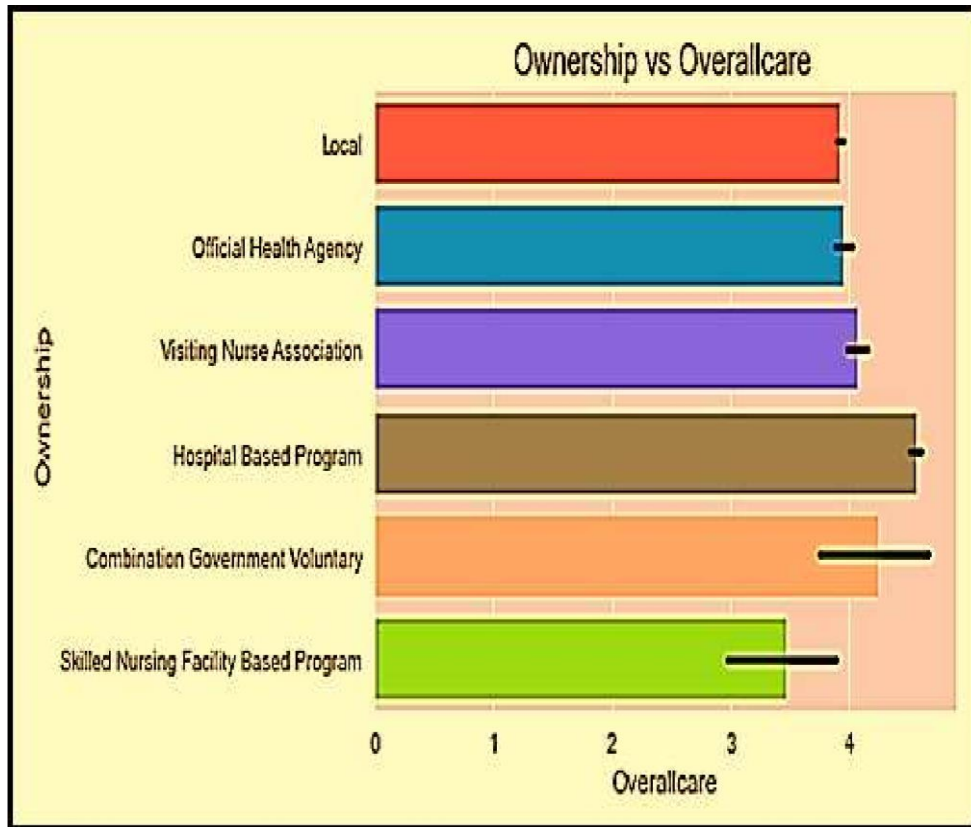
**Figure 3.23: ROC**

Those ownerships with the specified characteristic of our model prediction get the highest score, as shown clearly in Figure 3.21, which determines the factor of ranking-based service recommendations. Similarly, GB and RF both produced results with an accuracy higher than 99% for these datasets, but the others produced results with an accuracy lower than that (see Tables 3.13–3.16 and Figures 3.13–3.20). Consequently, these two models are considered top-tier models for using opinion

mining to forecast home healthcare services. The significance of characteristics in assessing model correctness is heavily influenced by patient judgment, as shown in Figure 3.21. Professionalism and patient safety are highly valued by patients, according to this survey, even more so than other qualities like communication and overall care. Figures 3.22 and 3.23 demonstrate the results of the model comparison and assessment, which corroborated the same by producing a ROC score.

Now that we have the criteria in hand, we can use them to identify the different home healthcare service ownership options. Patients may get services from a variety of ownership types, including local, official health agencies, visiting nurse associations, hospital-based programs, combination government voluntary programs, and staff nursing facility-based programs. These HHCS are assessed using a wide range of characteristics; however, in order to conduct our study, we have selected just the aspects that are highly dependent using our feature selection approach. Among the chosen characteristics are competence, safety, and open lines of communication.





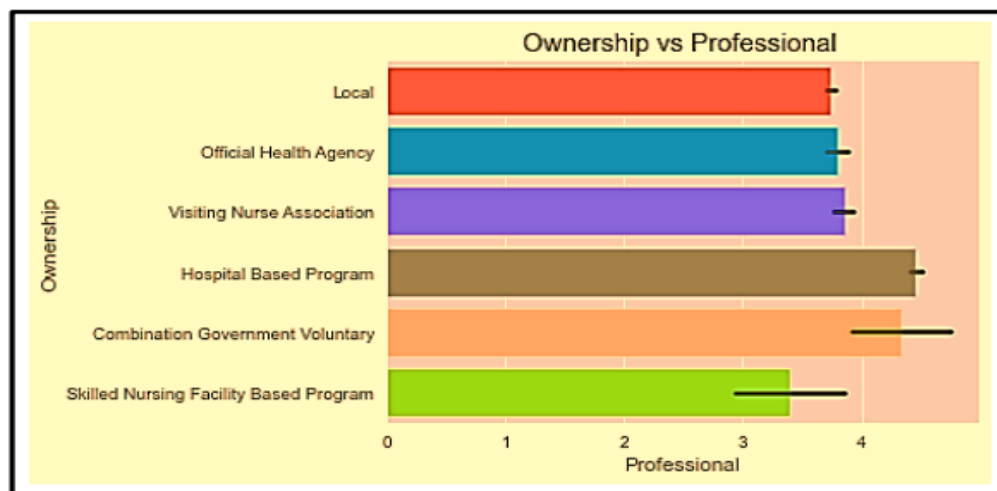


Figure 3.24: Features based on HHCS

Figure 3.24 provides a visual depiction of the link between home healthcare service ownership and the characteristics chosen according to patient satisfaction ranking score. When comparing all of the features offered by the agency, the diagram clearly shows that patients prefer hospital-based services, followed by combination government voluntary and visiting nurse associations, and other agencies come in third class. Both the ability to make effective recommendations and the ability to score them are essential features of a successful prediction model for recommendations. A recommendation system based on several ranking criteria for a certain dataset needs to take the important attributes into account. On this page, we provide a rating based on the ownership of different HHCS and attempt to forecast the greatest characteristics.

When compared to other ML models for feature ranking, Boosting approaches come out on top. In comparison to previous research, both the RF and GB models achieved scores of 99.868 and 99.824 respectively. The expert care and safety throughout treatment are the most important aspects of HHCS, according to patients, who suggest it to others. You may also find a ranking-based HHCS suggestion that takes ownership type into account when choosing features.

SENTIMENT ANALYSIS AND PERFORMANCE COMPARISON FOR EVALUATING PATIENT OPINIONS ON DRUG SATISFACTION

4.1 Sentiment Analysis for Evaluating The Patient Opinions on Drug Satisfaction

Experts and patients alike may benefit greatly from the drug recommendation framework's ability to shed light on medications for certain medical issues. A generic system for benefit-based medication recommendation is the subject of the planned study. Through the analysis of client feedback, these frameworks were able to deconstruct patients' emotions and provide personalized advice. The medication recommendation system uses sentiment analysis and function engineering to propose medications for certain illnesses according to patient ratings. The approach, results achieved by various models, and model validation are all included in the proposed paper.

To determine the tone of the data, Sentiment Analysis employs three distinct methods. A number of methods are based on machine learning: By splitting the dataset into training and test sets, this method applies the classification methodology to examine the opinions within the data. The purpose of training data is to study properties, while the purpose of test data is to evaluate the efficiency of algorithms. Using review and rating data from Amazon, TripAdvisor, Facebook, and Twitter, as well as other sources, Ducange and Fazzolari trained their classifiers in 2017 to detect feelings like positive, negative, and neutral. The findings show that the system can accurately classify 91.95% of TripAdvisor restaurant reviews and 93.11% of Amazon product ratings. Using a vocabulary-based method: Using this emotional terminology dictionary, one may find out the opinion words' polarity by matching them. This method employs strategies such as those based on manual analysis, dictionaries, and corpora.

4.1.1 Proposed Methodology

This research aims to examine a new approach of assessing perspectives on

medicines. Patients may compare medications based on their own evaluations and the suggestions of healthcare professionals for their individual situations. When doing sentiment analysis, many factors are taken into account to have a better understanding. Medication analysis in healthcare often makes use of text, ratings, and other quantitative data to create opinions. The purpose of the proposed model is to rank the dependent variables in terms of their significance to users' judgments of the medicine's efficacy for different medical conditions.

The suggested technique is schematically shown in Figure 4.1. The suggested approach is based on the following phases from UCI: This study made use of the drug proof dataset available in the Machine Learning Repository. Included are statistics on patient ratings for specific medications and associated health concerns, as well as a 10-star rating that stands for overall patient happiness. Medication attributes might include user preferences, ratings, names, conditions, dates of use, and more. We undertake feature extraction so that we can understand the scores and make accurate predictions. Data is partitioned into training and testing sets after preprocessing. The data is then prepared for further processing.

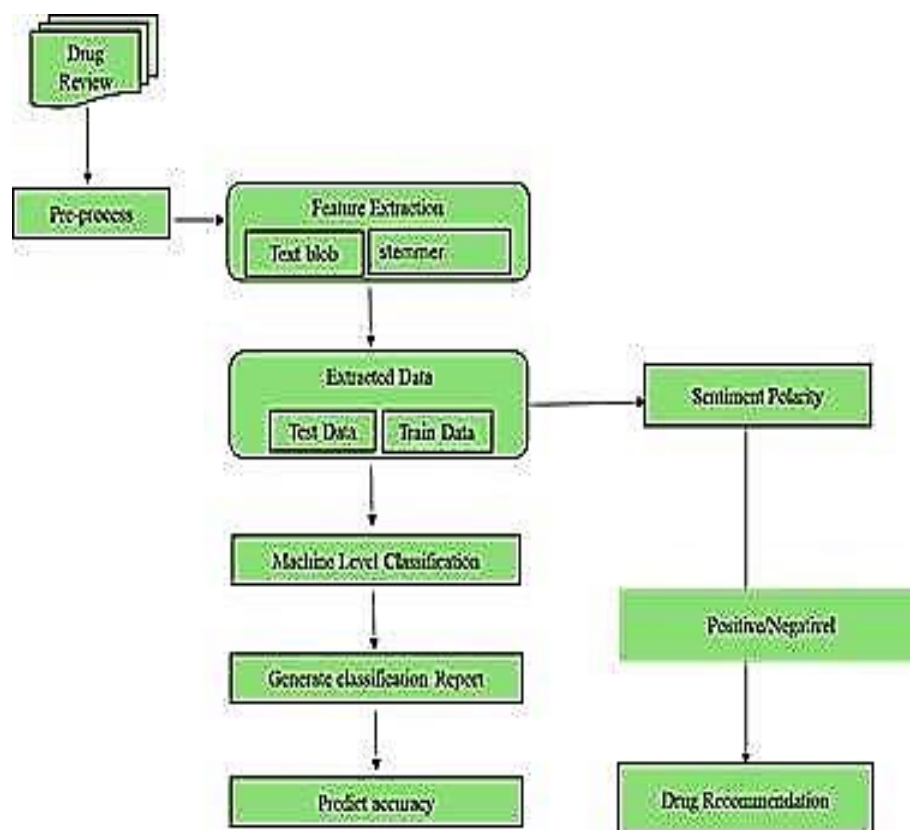


Figure 4.1 Proposed Methodology

Steps involved in proposed methodology:

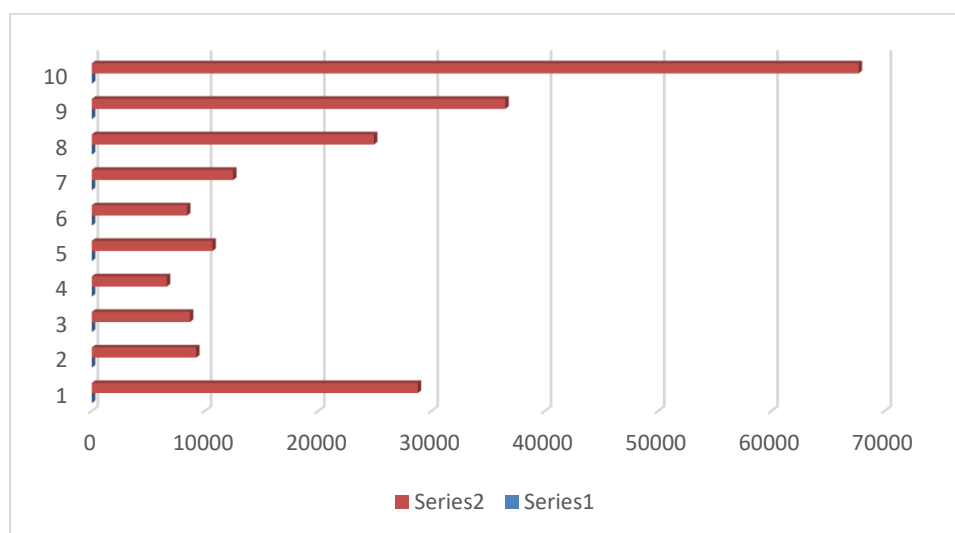
- 1) Begin
- 2) Import data from UCI Machine Learning Repository using Python
- 3) Apply stemmer and text blob to input data
- 4) Label encoder for changing categorical to numerical data for drug name and condition
- 5) Create sentiment polarity for cleaned data using Textblob
- 6) Divide the dataset into train and test datasets in a 70:30 ratio.
- 7) Implement ML Classifiers
- 8) Estimate the Performance Measure.
- 9) Use Cross-Validation to ensure model accuracy.
- 10) Drug Recommendation with sentiment polarity and positive score.
- 11) End

Data Pre-processing

A mood prediction classifier was developed using a variety of machine learning classification methods. As a result, when developing the recommendation system, we standardize the usable amount by applying criteria. Images 4.2 and Table 4.1 show the medication's 10-star ratings. The overall count of reviews was reduced after data pre-processing, from 215063 to 213869.

Table 4.1: User Ratings (1-10 Scale)

1	2	3	4	5	6	7	8	9	10
28769	9203	8662	6622	10650	8403	12470	24909	38499	67682

**Figure 4.2: User Ratings (1-10)**

A good sentiment is indicated by a rating of 5 or more, whilst a negative feeling is indicated by a rating of less than 5, depending on the polarity of the reviews. Figure 4.3 and Table 4.2 illustrate this below.

Table 4.2: No. of user ratings after conversion

No. of user ratings	
Positive	161491
Negative	53572

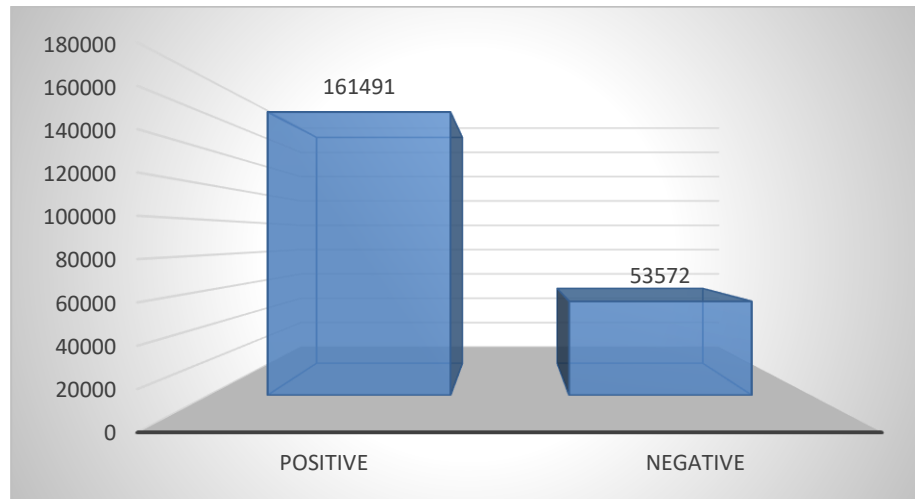


Figure 4.3: User Ratings (1/0)

Feature Extraction

Using a label encoder, we may transform the drug's name and the medical condition that it is prescribed for into numerical data. Following that, the datasets were subjected to sentiment polarity analysis. Independent Component Analysis and Principal Component Analysis are both used in this investigation.

Classification Approaches

- **LG Boost**

The LG Boost method is a fast-distributed gradient boosting framework that uses a decision tree technique to decrease memory use and achieves great performance. Ranking and classification issues are common uses for these methods.

- **XG Boost**

XG Boost builds decision trees in a sequential fashion. As with other independent variables, weights play a significant role in XG Boost. The decision tree is then fed these weights in order to make predictions. The variables are put into the second decision tree with their weights enhanced in the event that the prediction is erroneous.

A more accurate and reliable model is produced by combining these separate classifiers and predictors. Issues with ranking, classification, regression, and user-defined predictions get the best results when these techniques are used.

- **Cat boost**

To improve decision trees' gradient, the engineers and researchers at Yandex developed Cat boost. Several additional corporate processes rely on it, including searching, recommendation systems, personal assistants, self-driving vehicles, weather forecasting, and CERN, Cloud Flare, and Careem Taxi. Here is one way to describe the Cat boost parameters:

Cat Boost Classifier (iterations = 10000, learning rate = 0.5)

4.1.2 Results and Discussions

Python was necessary for data extraction from the UCI Machine Learning repository. Using label encoders, we can convert drug names and conditions from categorical to numerical data. Preprocessing the data is the next stage, as shown in Table 4.2. The 10-star ratings for the drugs may be shown in Figure 4.1. Due to data pre-processing, the overall count of reviews was reduced from 215063 to 213869. In order to create test and train sets, these datasets are divided 70:30. The data is sorted into positive and negative categories, as shown in Figure 4.10. If the user rating is more than 5, we consider it favorable; if it's less than 5, we consider it negative. As can be shown in tables 4.3 to 4.5, the boosting strategy is then used by the machine learning classifiers to predict the model's accuracy. The accuracy of the model was evaluated using a number of metrics, such as the F1-measure, recall, and precision.

Table 4.3 and Figure 4.4 display the results of the LGBM model classification, as indicated in Figure 4.10. This process involves cross-validating many machine learning models using accuracy metrics in order to identify the best model performance.

An accuracy of 0.91 is predicted for a positive mood and an accuracy of 0.85 for a negative mood. The expected memory values for good and negative emotions are 0.96 and 0.73, respectively. The LGBM was able to reach an accuracy of 0.91249.

LG Boost Accuracy

The model's accuracy is: 0.9124977946104331

The Matrix of Confusion is

Table 4.3: Classification report on LG Boost

LG Boost Accuracy: 0.9124977946104331			
Sentiments	Precision	Recall	F1-Score
Negative	0.85	0.73	0.79
Positive	0.91	0.96	0.94

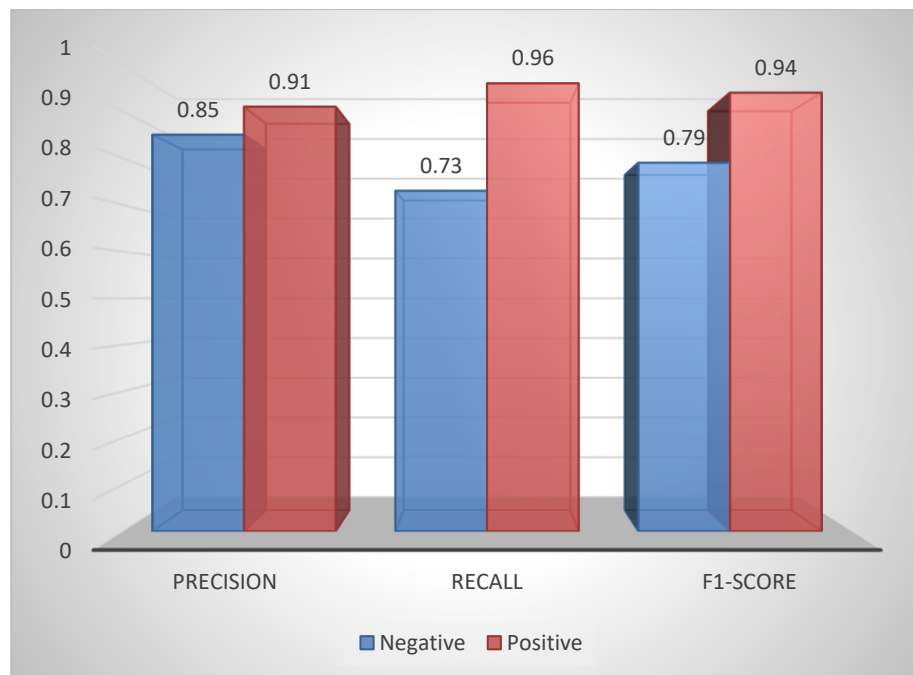


Figure 4.4: Performance Metrics Comparison for Negative and Positive Classes on LG Boost

The outcomes of the XG Boost model's classification are shown in Figure 4.5 and Table 4.4 below. The accuracy for predicting negative opinions is 0.68 while the accuracy for predicting positive opinions is 0.81. As a result of this procedure, good sentiment has a recall value of 0.95 and negative sentiment a recall value of 0.34. An accuracy of 0.79521 was achieved by the XGB.

XGB Accuracy

The Accuracy of the model is: 0.7952182790168483

The confusion Matrix is

Table 4.4: Classification report of XG Boost

XG Boost Accuracy: 0.8052182790168483			
Sentiments	Precision	Recall	F1-Score
Negative	0.68	0.34	0.46
Positive	0.81	0.95	0.87

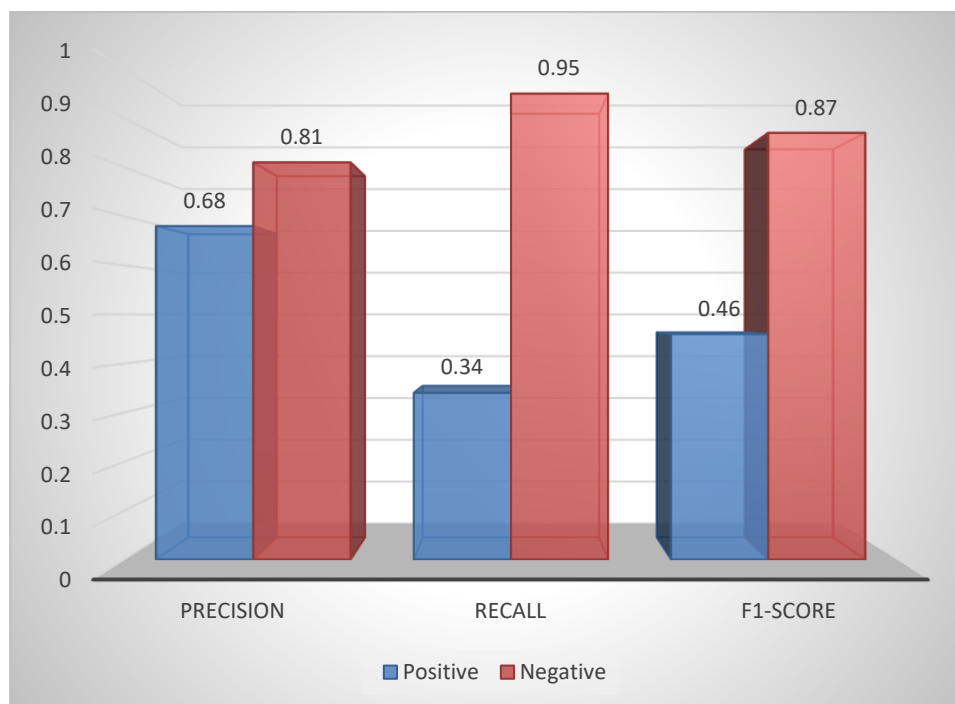


Figure 4.5: Performance Metrics Comparison for Negative and Positive Classes of XG Boost

Below, in Table 4.5 and Figure 4.6, you can see the Cat Boost model's classification report. With a favorable view, the projected accuracy is 0.82 and a negative one is 0.74. The positive sentiment recall score is 0.95 while the negative sentiment recall value is 0.74.

Cat Boost Accuracy

The Accuracy of the model is: 0.8942971587101198

The confusion Matrix is

Table 4.5: Classification report of Cat Boost

Cat Boost Accuracy: 0.8942971587101198			
Sentiments	Precision	Recall	F1-Score
Negative	0.82	0.74	0.78
Positive	0.92	0.95	0.93

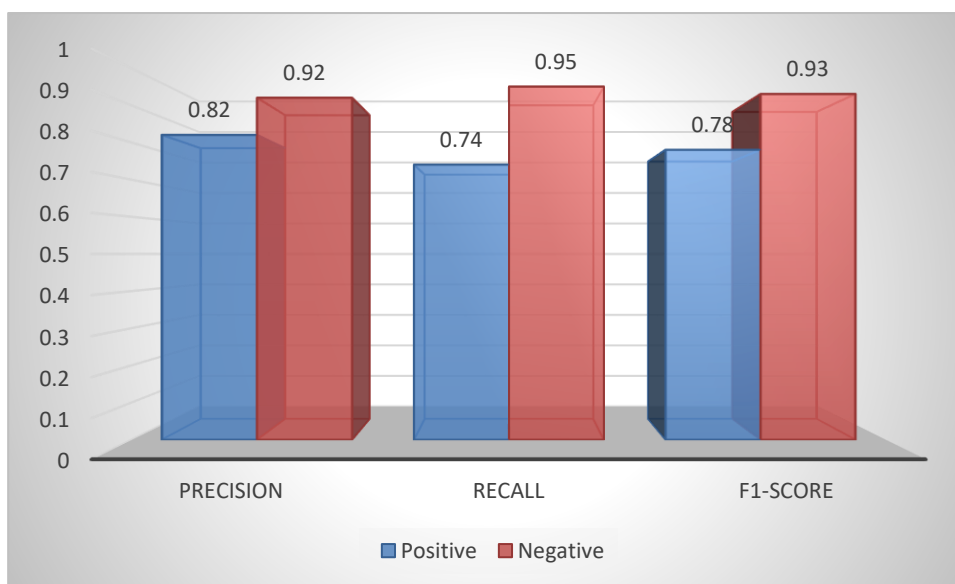


Figure 4.6: Performance Metrics Comparison for Negative and Positive Classes of Cat Boost

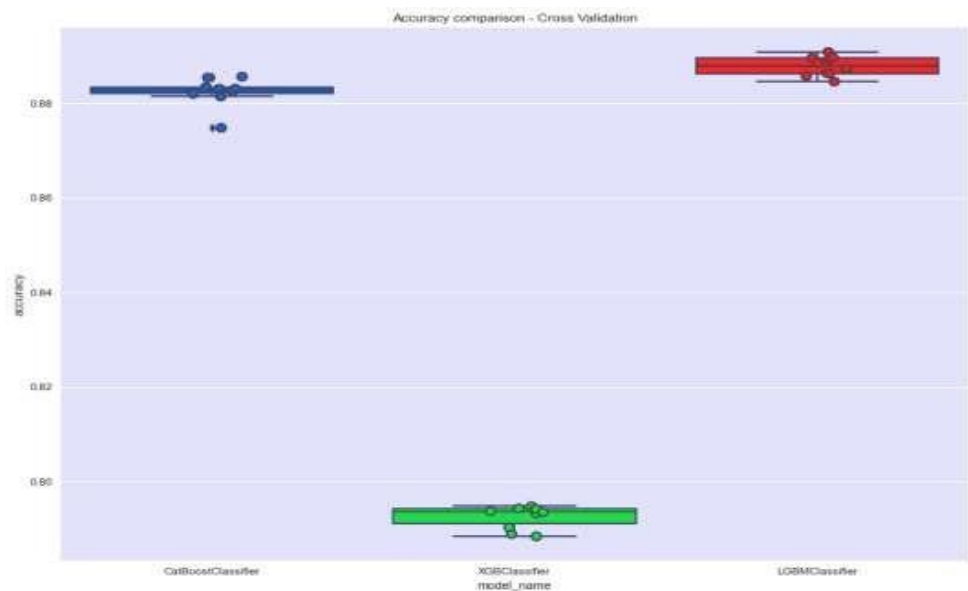


Figure 4.7: Accuracy

Figure 4.8 shows the top 20 situations in the dataset with the highest number of 3 medicine reviews, based on the sentiment polarity score computed from text blob.

4.1.3 Model Validation

Models' classification performance as shown in Table 4.6 and Figure 4.9 is influenced by the Boosting strategies used. The capacity to classify sentiment is assessed using indicators of recall, accuracy, and F1-measure. The purpose of cross validation is to compare the performance of several models. Boosted sentiment classifier model results are shown in Tables 4.4-4.5 for each classifier model.

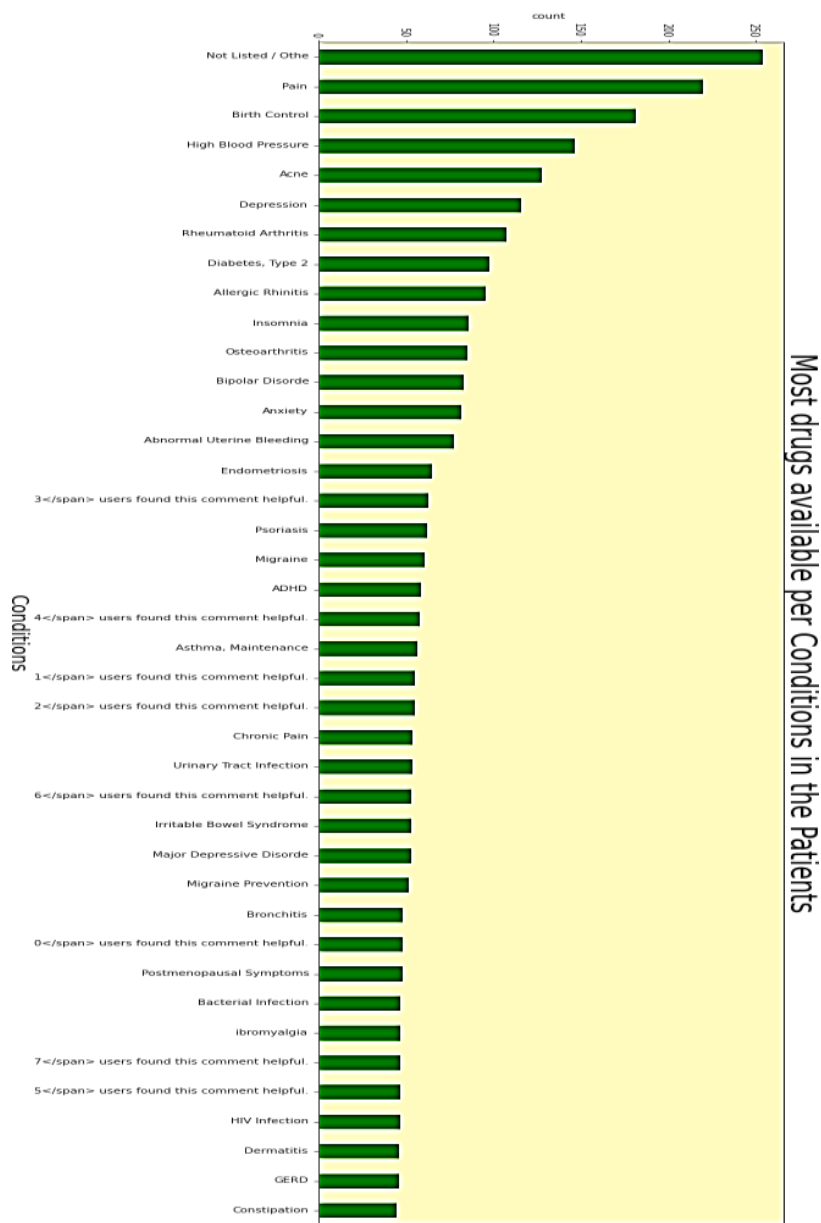


Figure 4.8: Top drugs available per conditions in patients

The obtained model prediction scores are validated using a receiver operating characteristic (ROC) curve. Classifiers that are closer to the top-left corner of the screen indicate better model performance. In Figure 4.7, we can see that the different emotion classification models have their color coding reversed. Compared to other models, our calculations demonstrate that LGboost performs better. Its model accuracy is 0.91 and its ROC accuracy is 0.85, both of which are somewhat better than earlier studies. To analyze the data, we employ sentiment evaluations, both positive and negative. To treat a certain illness, a drug has to have a high grade.

Model validation using 10-fold cross-validation and ROC-AUC score validation is shown in Figure 4.7.

Table 4.6: Prediction metric scores for sentiment boosting classifier models

Classifier Model	F1	Model Accuracy	Precision	Recall	AUC-ROC
LG Boost	79	91	85	73	85
XG Boost	46	80	68	34	64
Cat Boost	78	89	82	74	85

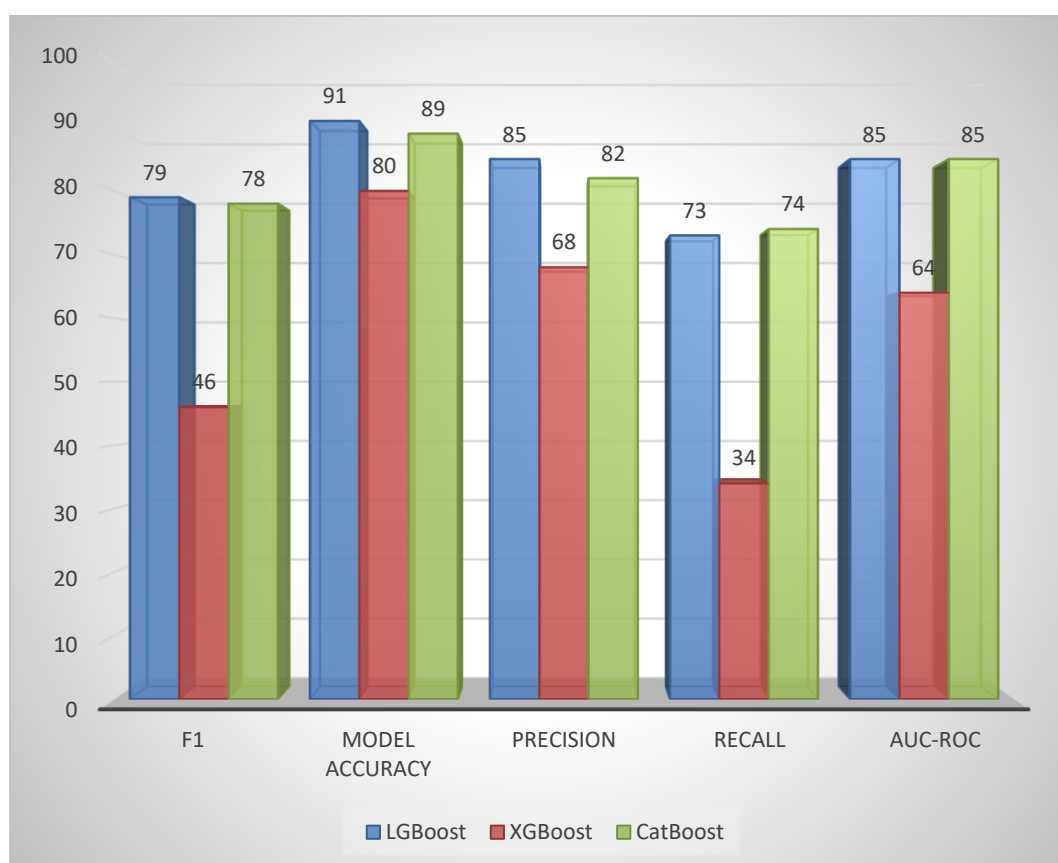


Figure 4.9: Prediction Metric Score

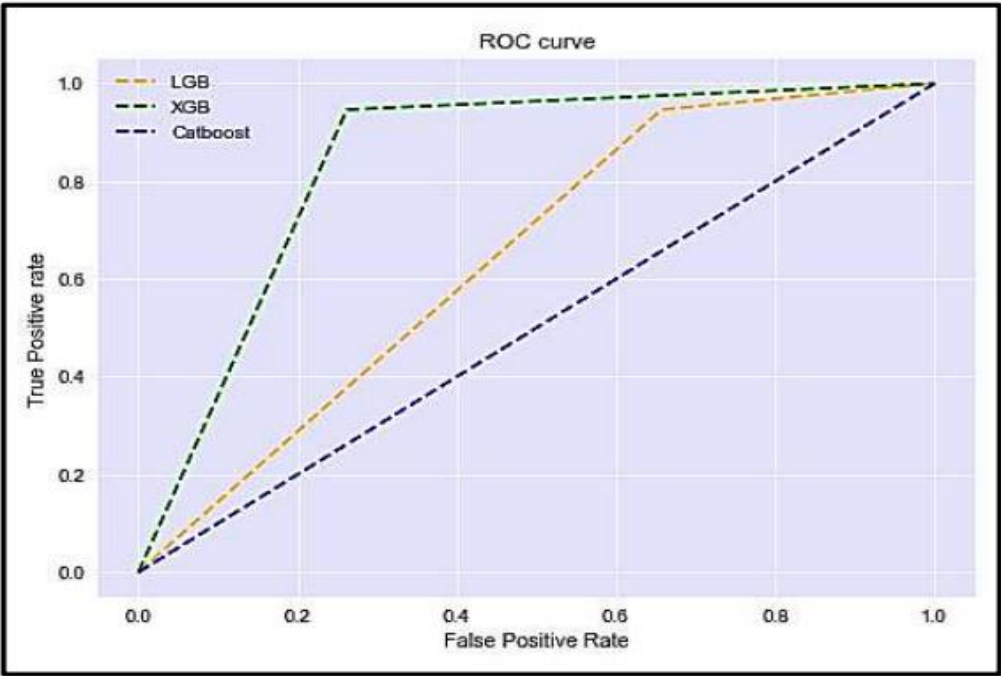


Figure 4.10: Comparison - ROC-AUC curve

Medications recommended for each disease are shown in Figure 4.11. Medication recommendations are based on dividing the sum of positive numbers by the corresponding condition's emotion polarity score.

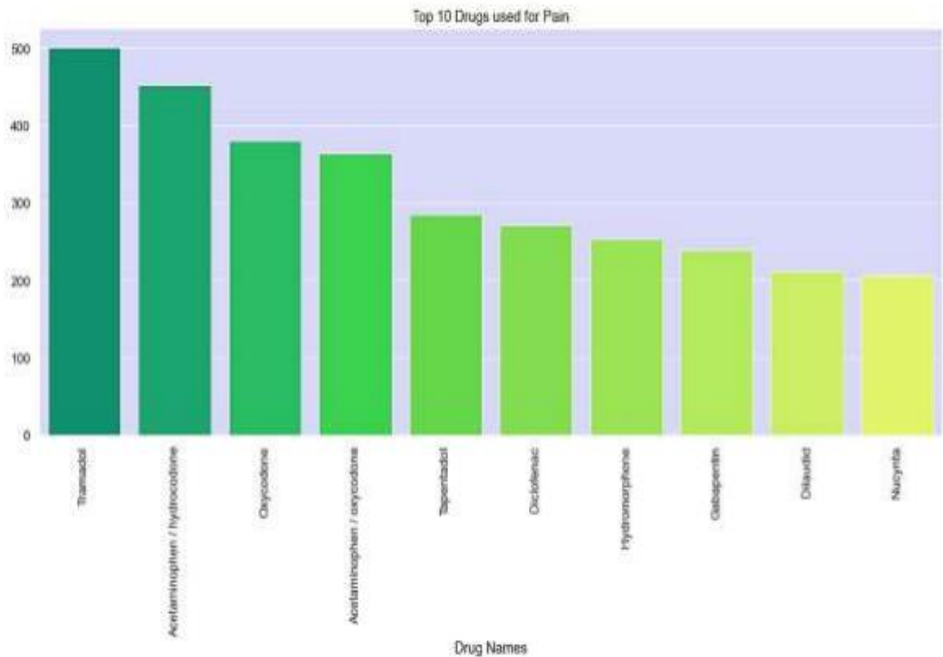


Figure 4.11: Top 10 drugs

Though the recommendation architecture for real-time applications is not shown, the results of each classifier demonstrate the model's effectiveness based on assumptions from user ratings. If you want greater understanding from your data, you need make an additional enhancement. When making a recommendation for a system, the only factors taken into account are the review's polarity score and the total number of users. In this work, we sought to provide a new approach to medication recommendation using categorization metrics and review sentiment.

Predicting patients' thoughts on medication reviews is the aim of this suggested study. Using the Light GBM, XG Boost, and Cat Boost models, the projected accuracies scores for this proposed study are 0.91, 0.79, and 0.89, respectively. Due to its superior accuracy, the LGBM classifier outperforms all of its competitors. Since accuracy is only one metric, the categorization report and model cross validation provide a richer picture of the model.

4.2 Performance Comparison

In order to find ways to improve healthcare and make the end user happy, it is helpful to get sentiment classifications like positive, negative, and neutral. The healthcare industry is able to maximize customer happiness via the optimization of market strategies, according to the outcomes of these analyses.

The use of opinion mining and sentiment analysis to investigate emotional and behavioral patterns has recently been the subject of a great deal of study. Online forums have become the go-to place for individuals to express themselves in this technologically advanced day.

It has also been noted that individuals like to utilize these online forums to express themselves emotionally, intellectually, and internationally. When determining their market worth, top management relies heavily on the data supplied by end customers. By classifying the review data according to its polarity, we can determine if the sentiment is favorable, negative, or neutral in these online reviews.

The most important part of assessing user sentiment is pre-processing, which is necessary before analyzing any data, whether structured or unstructured. After the sentiment analysis, the relevant product, object, or domain is suggested according to

the sentiment score and polarity. Retrieving health-related data from an internet source and recommending the best data based on user demand is the basic goal of health recommendation. A user's specified preferences and various filter techniques are used to present a collection of things in response to an online search, with the goal of providing a well-refined suggestion for the user's criteria. One of the greatest ways to get the information you need is to use a recommendation system.

A system-based intelligent approach, it provides consumers with a wealth of information in response to their questions. It uses a variety of methodologies in different sectors to deliver people personalized results. In order to provide recommendations, modern systems use a plethora of machine learning algorithmic techniques. We provide a comparative study here by reading and comparing a number of research articles on recommendation systems that use online forums and sentiment analysis.

4.2.1 Comparison Analysis

In order to maintain the high standard of our work, we used the following criteria to choose papers: Articles chosen beginning in 2015. 1. The document has to be suitable for the search, 2. The language of the paper should be English. 3. The paper's results and explanations should be in a logical order. Third, a reputable journal should be cited in the paper. 4. The well-respected digital resources used to compile these publications include IEEE Explore, Springer, ResearchGate, Science Direct, Pub Med, and ACM.

4.2.2 Patient Opinion Mining in Healthcare-Related Recommendation

Whether it's directly from their mouths or their review comments, we often find it helpful to hear other people's thoughts while making decisions in our daily lives. If we need a present but aren't sure what to get, we may always ask the store clerk for advice. We may see the shopkeeper's recommendations or other products displayed before making a purchase. In a similar vein, people often go to internet forums for suggestions.

Inputting the user's opinion, the recommender system will then utilize different filtering techniques to extract the relevant information and deliver an

efficient and necessary message as an output.

Its widespread use has contributed to its recent surge in popularity. The health care industry is in dire need of this recommendation system due to the sheer number of areas that need to be assessed and suggested according to consumer preferences.

Listed below are a few healthcare recommendations for hospitals, healthcare services, and medication ideas that have been shown to increase patient satisfaction. Many areas still need to be investigated.

All of the aforementioned healthcare-related topics were our goal in producing this work. We used machine learning methods to uncover previously unknown facts after collecting data from internet forums.

The use of opinion mining and sentiment analysis in hospital suggestions enables the end user to make an efficient decision according to their requirements. In this section, we compiled the reviews of various healthcare facilities. We discovered the model's correctness and if it delivers a decent score compared to current ones by using natural language processing and machine learning approaches to assess the sentiment polarity between ratings.

S. Kethavarh used decision trees and Naïve Bayes classifiers in 2015 using the Tweeter dataset, and their model achieved an accuracy of 80.08% and 81.21%, respectively, for decision trees. Using machine learning classifiers, Raj.S. Suresh also suggested a model for the Twitter dataset in 2015. His method yielded 88% for SVM and 76% for Naïve Bayes.

Mrs. R. Nancy Deborah reviewed products online in 2021 using ML classifiers. Their model achieved an accuracy of 80% in LR, 80% in RF, and 80% in SVM, according to the current model's contributions. Using machine learning classifiers and the ensemble approach, our proposed model analyzes Google reviews to suggest hospitals. Its accuracy was 93% for SVM, 85% for LR, 81% for KNN, and 95% for RF. It was tested using several hospital reviews and found to be the most accurate model.

Advantages

When compared to the performance and accuracy of individual models, an ensemble's predictions are superior. The ensemble method lessens the dispersion of both model performance and predictions.

Disadvantages

Due to the difficulty in predicting and explaining ensemble model output, the ensemble technique is not as interpretable as other methods. Not only is ensemble learning challenging, but incorrect selection may result in worse prediction accuracy compared to individual models.

Our suggested research takes into account user reviews and ratings of various healthcare services based on a variety of criteria. The comparison between the current models and the suggested models is shown in Table 4.7 below.

Table 4.7: Comparison on healthcare service recommendation based on accuracy

Model	Accuracy	Precision	Recall	F-Score
Deep learning	79	0.79	0.8	0.79
CNN	82.6	0.85	0.84	0.84
RF	97.75	95	94	94
Boosting Algorithm	99.82	1	0.99	0.99

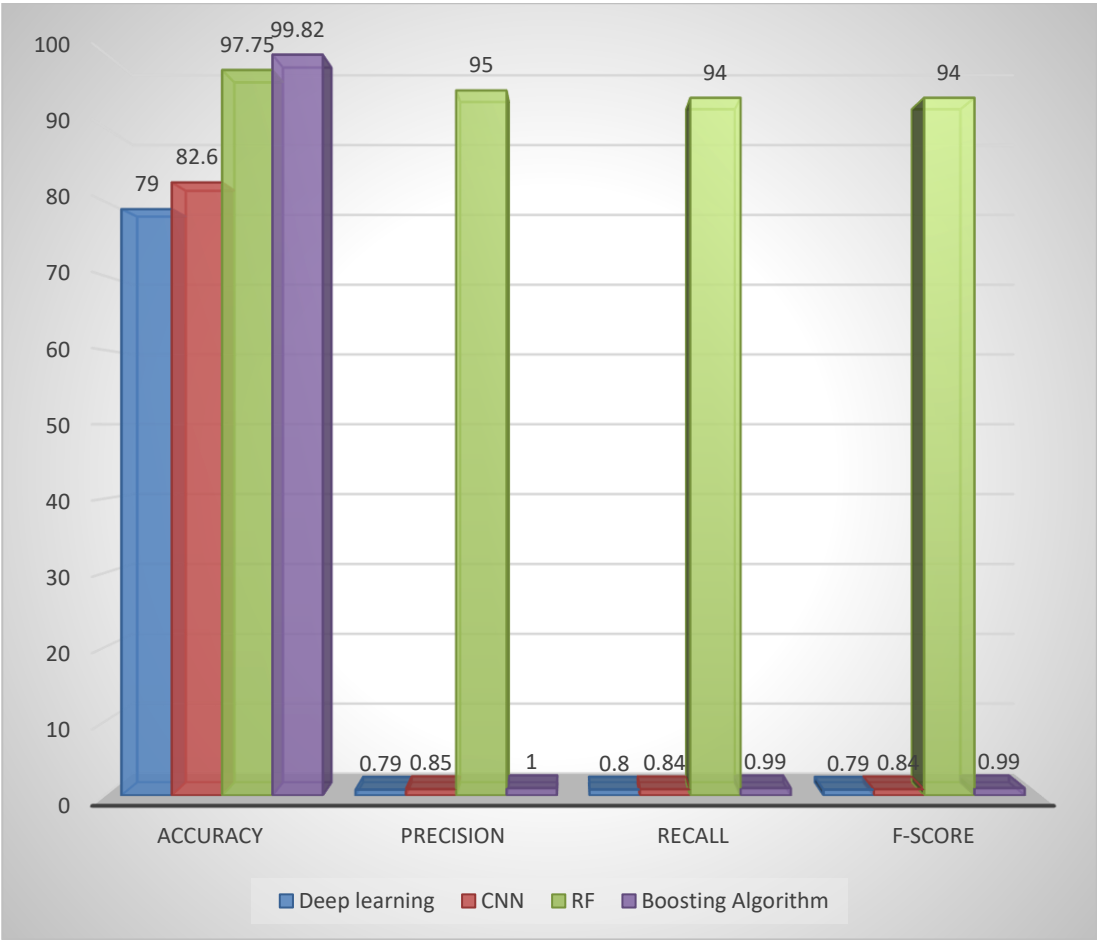


Figure 4.12: Graph showing Comparison on healthcare service recommendation based on accuracy

According to Table 4.7 and Figure 4.12, our suggested approach outperforms the current models in terms of accuracy. One potential area of recommendation system development recently is medication recommendations. The analysis of this user data has been the subject of several types of study.

Because many people want to hear about potential adverse effects and feedback from peers who have tried alternative medicine before taking a pill, this is a difficult topic to navigate.

The analysis of these drug review datasets has included a number of different methods. In this section we compared our technique to the other research endeavors in this field, and table 4.8 below shows the results. In terms of reaching the intended model's correctness, our model is found to be superior to other current models.

Table 4.8: Prediction metric scores for sentiment boosting classifier models

Domain	Methods	Accuracy	Precision	Recall	F-Score
AskPatient.com	Ensemble Approach	78.2	56	66	-
AskPatient.com	International Model	78	56	66	-
Online Forums	LR	80	-	-	-
	GNB	72			
	RF	80			
	SVM	80			
	XGB	87.5			
Drugs.com	D.T.	78	86	81	84
	Cat Boost	88	91	92	92
Drugs.com	LGBM,	91	85	73	79
	Cat Boost	89	82	74	78

Advantages of Opinion Mining and Sentiment Analysis

By using sentiment analysis and patient opinion mining, hospitals may get a deeper understanding of patients' actions and opinions. Because of this, hospitals are able to better understand their patients and, with that knowledge, provide better care. One benefit is the ability to establish a strong relationship with patients. Another is that it can increase patient satisfaction and service quality. Still another is that it can monitor the hospital's brand image. Lastly, it can help with social listening and the detection of negative comments about the hospital's branding and service.

Dis-Advantages of Opinion Mining and Sentiment Analysis

When used to noisy data from many domains, Opinion Mining and Sentiment analysis methodologies provide less efficient results. This is because the variations in people's expressions across these fields are insufficient to warrant processing them efficiently.

FINDINGS, CONCLUSION, SUGGESTIONS AND FUTURE SCOPE OF THE STUDY

5.1 Findings of The Study

1. Increased Predictive Accuracy in Patient Satisfaction

The study found that machine learning classifiers significantly enhance the accuracy of predicting patient satisfaction levels. By analyzing historical patient data and feedback, algorithms such as decision trees, support vector machines, and neural networks were able to identify patterns and predictors of satisfaction with high precision. This predictive capability allows healthcare providers to proactively address potential issues.

2. Improved Service Quality Metrics

The integration of big data analytics enabled the identification of key service quality metrics most closely associated with patient satisfaction. Attributes such as waiting time, staff behavior, and clarity of communication were highlighted as critical areas for improvement. The insights derived from data analysis guided strategic decision-making to enhance these metrics effectively.

3. Personalization of Patient Care

Machine learning models demonstrated the ability to support the personalization of care by segmenting patients into groups based on their needs, preferences, and historical behavior. Personalized recommendations improved the patient experience, ensuring tailored healthcare delivery and fostering trust in the system.

4. Real-Time Feedback Mechanisms

The use of big data analytics facilitated the development of real-time feedback mechanisms. By integrating patient feedback systems with analytics platforms,

healthcare organizations could monitor satisfaction levels dynamically, enabling rapid responses to emerging concerns and improving patient retention rates.

5. Cost-Effective Resource Allocation

The study revealed that data-driven insights from machine learning models helped optimize resource allocation. Predictive models identified areas where operational efficiency could be improved without compromising service quality, such as staffing schedules, inventory management, and appointment systems.

6. Enhanced Decision Support Systems

Big data analytics strengthened decision support systems for healthcare administrators. Predictive models informed policy and process changes by presenting actionable insights based on extensive data analysis. This capability resulted in strategic improvements in service delivery and operational management.

7. Reduction in Patient Complaints

An observable outcome of implementing machine learning and analytics was a reduction in patient complaints. By addressing key pain points identified through predictive analytics, healthcare providers succeeded in resolving common grievances, enhancing the overall patient experience.

8. Scalability and Future Readiness

The study highlighted that the integration of machine learning classifiers and big data analytics is scalable, making it suitable for healthcare institutions of varying sizes. Furthermore, it positions these organizations to adopt future innovations in artificial intelligence and data-driven healthcare solutions.

5.2 Conclusion

Review, survey, comment, and data mining extraction is a subset of data mining that aims to find useful information in large datasets. This information is useful for understanding the sentiments expressed in the comments and for assessing the service's ability to meet the needs of its users. These online forum conversations are now very important for determining the domain's market worth from the user's

point of view. A lot of work goes into deciphering these viewpoints and drawing useful conclusions from the data they provide. According to their needs, the user is receiving the suggestion. One useful tool that has recently surfaced to aid users in discovering and evaluating things of interest is sentiment analysis recommender systems. In order to help customers find products that are a good fit for their tastes or requirements, these systems utilize a variety of methods. Hybrid systems are seeing a rise in the utilization of data mining approaches to enhance recommendations, even if well-known CF-based algorithms still provide relevant, individualized results across many disciplines. Optimal suggestions in challenging domains were previously generated by effective apps and standalone recommenders.

Additionally, apps have shifted away from advising what to consume due to changes in machine learning classifier algorithms, which have changed the kinds of recommendations. Although recommender systems were once seen as a fleeting trend, they have since shown to be a practical and effective tool with many potential uses. These algorithms can and will play a significant role in sentiment analysis-based recommendation systems.

Technology has the ability to revolutionize healthcare service delivery, as shown in research that aimed to improve patient happiness and service quality by integrating machine learning classifiers and big data analytics. Healthcare professionals may better address the needs of their patients by using machine learning algorithms to sift through mountains of patient data in search of patterns, outcomes, and opportunities for personalized treatment. By analysing various datasets, finding previously unknown insights, and facilitating data-driven decision-making to enhance healthcare services, big data analytics greatly enhances these skills. Response times, treatment results, and patient satisfaction are some of the important performance measures that are greatly improved by using sophisticated computational methodologies, as shown in this study. Predicting patient satisfaction levels, identifying service bottlenecks, and making targeted adjustments are all within the capabilities of machine learning classifiers, especially when linked with big data analytics. With the help of these resources, healthcare organizations may move away from reactive service models and toward proactive ones, encouraging a mindset of constant improvement.

In addition, the research emphasizes that healthcare practitioners and patients alike may benefit from patient-centric methods that are backed by technology improvements since they make the process easier and faster. It stresses the significance of data privacy and ethical concerns in using these technologies to keep faith and stay in line with regulations. Finally, this study lays forth a solid plan for healthcare organizations to use big data analytics and machine learning to improve service quality and patient happiness. In order to make the most of these instruments, it stresses the need of maintaining investments in infrastructure, training, and technology. Healthcare service delivery and patient outcomes might be further optimized with the use of more advanced algorithms, real-time data, and natural language processing for patient feedback.

A new SCSP Ensemble Model for Health Care Opinion System Analysis is presented in this study. Using a machine learning strategy to forecast superior models in data analysis, the classification models categorize patients' emotions as positive, negative, or neutral. Then, this study presents a new feature selection approach that uses patient satisfaction data to find the most important aspect of home health care services. In this, we looked at the many parts that make up health care service superiority evaluations using various metrics. The next section of this study proposal details a new approach to sentiment analysis that uses symptoms to assess patients' levels of medication satisfaction via the suggestion technique. The next step is to compare the suggested methods' performance to that of the current systems, and finally, to give the reasoning behind the suggested methods.

By evaluating healthcare-related social media evaluations, our suggested framework achieves an AUC accuracy of 95% and an overall accuracy of 93%. Based on good ratings, it gives a recommendation % for the healthcare industry. The story concludes with a summary of the planned effort, including its unique aspects and its contributions. The paper lays out the suggested methods for future study as well as the outstanding problems in the health care arena.

5.3 Suggestions of The Study

1. **Implementation of Predictive Models:** Machine learning classifiers can help healthcare organizations anticipate patient needs and solve problems before they

ever start. Improved service delivery, process simplification, and resource prioritization are all possible with the use of predictive models.

2. **Integration of Big Data Analytics:** Big data analytics solutions should be purchased by healthcare facilities so that they may examine operational data, medical records, and patient feedback. Insights gleaned from these platforms may be used to improve service quality and patient outcomes.
3. **Personalized Care Plans:** Healthcare providers may enhance patient happiness and build confidence in healthcare services by using machine learning insights to create customized treatment plans that are suited to individual requirements.
4. **Focus on Staff Training:** Machine learning models provide data-driven insights that healthcare personnel should be educated to comprehend and use. The seamless incorporation of technology into routine tasks may thus be guaranteed.
5. **Ethical and Privacy Considerations:** Organizations should establish strong data governance rules to guarantee the ethical handling of patient data and compliance with privacy standards like GDPR or HIPAA. The success of efforts powered by technology depends on maintaining the confidence of patients.
6. **Continuous Monitoring and Feedback:** Incorporate real-time data analytics into existing systems to continuously assess service quality and patient satisfaction. Healthcare may be made more dynamic and responsive by using patient input to enhance services.
7. **Collaboration with Technology Providers:** To get cutting-edge resources and make sure machine learning and big data analytics are seamlessly integrated into healthcare institutions' systems, they should work with technology providers.
8. **Utilization of Real-Time Analytics:** Address urgent patient problems, track staff performance, and respond quickly in emergency circumstances with the use of real-time statistics.
9. **Targeted Improvements Based on Insights:** Reducing wait times, boosting communication, and upgrading the physical environment of healthcare facilities

are some of the areas where data analytics have shown to have the biggest influence on patient satisfaction.

10. Future Research and Innovation: The use of natural language processing to evaluate patient input and the development of more sophisticated machine learning algorithms merit more investigation. Improving patient engagement and satisfaction may also be achieved by investigating emerging technologies like chatbots driven by artificial intelligence.

With these suggestions in place, healthcare businesses can design service delivery models that put patients first, while also making use of big data analytics and machine learning to boost satisfaction and improve results.

5.4 Future Scope of The Study

Systematic evaluations of patient opinions in health care utilizing sentiment analysis are well-suited to the methods given here. Hotels, schools, malls, sports training facilities, supermarkets, restaurants, private bus transportation systems, domestic and international airlines, clothing stores, real estate, banking, service center for two-wheelers and four-wheelers, etc., can all benefit from the research approaches that have been suggested.

Future work may concentrate on making conversation bot and parsing mechanisms better able to handle larger data sets. For high-quality data collecting, which guarantees effective analysis and guaranteed results, multilingual sentiment classifiers may also be included into consumer evaluations and ratings.

REFERENCES

Chapter 1

- [1] **M. Bachir, M. Z. Fatima, L. Professor, M. Mezati, R. Azizi, and A. Djouhri**, "Enhancing Customer Satisfaction using Machine Learning," *Brazilian Journal of Technology*, vol. 7, no. 4, pp. 1–20, 2024.
- [2] **T. Sukmawati, T. Ramadania, and W. Pebrianti**, "Enhancing Patient Satisfaction by Healthcare Service Providers: A Systematic Literature Review," *Asian Journal of Economics, Business and Accounting*, vol. 24, no. 12, pp. 342–356, 2024.
- [3] **A. Prakash and R. Mohanty**, "Understanding service quality," *Production Planning & Control*, vol. 24, no. 12, pp. 1–16, 2012.
- [4] **A. N. K. Naik, S. Gantasala, and V. Gantasala**, "Service quality (servqual) and its effect on customer satisfaction in retailing," *European Journal of Social Sciences*, vol. 6, pp. 231–243, 2010.
- [5] **A. Lewis and V.-W. Mitchell**, "Defining and Measuring the Quality of Customer Service," *Marketing Intelligence & Planning*, vol. 8, no. 6, pp. 11–17, 1990.
- [6] **P. Hernon and D. Nitecki**, "Service Quality: A Concept Not Fully Explored," *Library Trends*, vol. 49, no. 4, pp. 1–10, 2001.
- [7] **J. Kenagy, D. Berwick, and M. Shore**, "Service Quality in Health Care," *JAMA: The Journal of the American Medical Association*, vol. 281, no. 17, pp. 661–665, 1999.
- [8] **M. Darzi, S. Islam, S. Khursheed, and S. Bhat**, "Service quality in the healthcare sector: a systematic review and meta-analysis," *LBS Journal of Management & Research*, vol. 21, no. 1, pp. 1–10, 2023.

- [9] **B. Prakash**, "Patient Satisfaction," *Journal of Cutaneous and Aesthetic Surgery*, vol. 3, no. 3, pp. 151–155, 2010.
- [10] **N. Garg, S. Kumar, and R. Mahesh**, "Patient Satisfaction Survey at a Tertiary Care Specialty Hospital," *International Journal of Research Foundation of Hospital and Health Care Administration*, vol. 2, no. 2, pp. 79–83, 2014.
- [11] **N. Elgendy and A. Elragal**, "Big Data Analytics: A Literature Review Paper," *Lecture Notes in Computer Science*, vol. 8557, no. 1, pp. 214–227, 2014.
- [12] **M. Memon, S. Soomro, A. Jumani, and M. Kartio**, "Big Data Analytics and Its Applications," *Annals of Emerging Technologies in Computing*, vol. 1, no. 1, pp. 1–10, 2017.

Chapter 2

- [1] **O. Alkhnbashi, R. Nassr, and M. Hammoudeh**, "Aspect-Based Sentiment Analysis of Patient Feedback Using Large Language Models," *Big Data and Cognitive Computing*, vol. 8, no. 12, pp. 2-29, 2024, doi: 10.3390/bdcc8120167.
- [2] **G. Okorie, Z. Egieya, U. Ikwue, C. Udeh, E. Adaga, O. DaraOjimba, and O. Oriekhoe**, "Leveraging Big Data for Personalized Marketing Campaigns: A Review," *International Journal of Management & Entrepreneurship Research*, vol. 6, no. 1, pp. 216-242, 2024, doi: 10.51594/ijmer.v6i1.778.
- [3] **L. Tanasescu, A. Vines, A.-R. Bologa, and O. Virgolici**, "Data Analytics for Optimizing and Predicting Employee Performance," *Applied Sciences*, vol. 14, no. 1, pp. 3254, 2024, doi: 10.3390/app14083254.
- [4] **A. Gupta**, "The Convergence of Big Data Analytics and CRM Practices: A Review," *International Journal of Computer Trends and Technology*, vol. 72, no. 7, pp. 74-82, 2024, doi: 10.14445/22312803/IJCTT-V72I7P109.
- [5] **M. Usman, M. Mujahid, F. Rustam, E. Soriano Flores, J. L. Mazón, I. De la Torre Díez, and I. Ashraf**, "Analyzing Patients Satisfaction Level for Medical Services Using Twitter Data," *PeerJ Computer Science*, vol. 10, no. 3, pp. 1697, 2024, doi: 10.7717/peerj-cs.1697.
- [6] **O. Amoo, F. Usman, E. Okafor, O. Akinrinola, and N. Ochuba**, "Strategies for Leveraging Big Data and Analytics for Business Development: A Comprehensive Review Across Sectors," *Computer Science & IT Research Journal*, vol. 5, no. 3, pp. 562-575, 2024, doi: 10.51594/csitrj.v5i3.861.
- [7] **N. Raju and P. Devi**, "A Comparative Analysis of Machine Learning Algorithms for Big Data Applications in Predictive Analytics," *International Journal of Scientific Research and Management (IJSRM)*, vol. 12, no. 10, pp. 1608-1630, 2024, doi: 10.18535/ijssrm/v12i10.ec09.
- [8] **C. Nwaimo, A. Adegbola, and M. Adegbola**, "Transforming Healthcare with Data Analytics: Predictive Models for Patient Outcomes," *GSC Biological and*

-
- Pharmaceutical Sciences*, vol. 6, no. 6, pp. 25-035, 2024, doi: 10.30574/gscbps.2024.27.3.0190.
- [9] **R. Chaudhary, H. Lohiya, and I. Journal**, "A Comprehensive Review on Machine Learning-Based Customer Satisfaction from Airline Tweets," vol. 5, no. 10, pp. 2582-6948, 2024.
- [10] **M. Ogrizović, D. Draskovic, and D. Bojić**, "Quality Assurance Strategies for Machine Learning Applications in Big Data Analytics: An Overview," *Journal of Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-01028-y.
- [11] **C. Udeh, O. Orieno, O. Daraojimba, N. Ndubuisi, and O. Oriekhoe**, "Big Data Analytics: A Review of Its Transformative Role in Modern Business Intelligence," *Computer Science & IT Research Journal*, vol. 5, no. 1, pp. 219-236, 2024, doi: 10.51594/csitrj.v5i1.718.
- [12] **M. B. Mahdjoub, F. Zohra, P. Laallam, M. Mezati, R. Azizi, and A. Djouhri**, "Enhancing Customer Satisfaction Using Machine Learning," *Brazilian Journal of Technology*, vol. 7, no. 4, pp. 1-20, 2024, doi: 10.38152/bjtv7n4-011.
- [13] **S. Fatima**, "Improving Healthcare Outcomes through Machine Learning: Applications and Challenges in Big Data Analytics," *International Journal of Research in Engineering Technology*, vol. 11, no. 12, pp. 2349-2819, 2024.
- [14] **Hanif and A. Jaafar**, "A Review: Application of Big Data Analytics in Airlines Industry," *Open International Journal of Informatics*, vol. 11, no. 2, pp. 196-209, 2023, doi: 10.11113/oiji2023.11n2.254.
- [15] **K. Barik, S. Misra, A. Ray, and A. Shukla**, "A Blockchain-Based Evaluation Approach to Analyse Customer Satisfaction Using AI Techniques," *Heliyon*, vol. 9, no. 1, pp. e16766, 2023, doi: 10.1016/j.heliyon.2023.e16766.
- [16] **M. R. Rashid, A. Nawaz, T. Ali, and G. Mustafa**, "Enhancing Healthcare Services Recommendation Through Sentiment Analysis," *Acta Universitatis*
-

- Sapientiae, Informatica*, vol. 15, no. 2, pp. 330-344, 2023, doi: 10.2478/ausi-2023-0020.
- [17] **N. Berros, F. Elmendili, Y. Filaly, and Y. EL BOUZEKRI EL IDRISI**, "Enhancing Digital Health Services with Big Data Analytics," *Big Data and Cognitive Computing*, vol. 7, no.2, pp. 64, 2023, doi: 10.3390/bdcc7020064.
- [18] **F. Del Giorgio Solfa and F. Simonato**, "Big Data Analytics in Healthcare: Exploring the Role of Machine Learning in Predicting Patient Outcomes and Improving Healthcare Delivery," *International Journal of Computations Information and Manufacturing (IJCIM)*, vol. 3, no. 1, pp.1-10. 2023, doi: 10.54489/ictim.v3i1.235.
- [19] **S. Rangineni, A. Bhanushali, M. Suryadevara, S. Venkata, and K. Peddireddy**, "A Review on Enhancing Data Quality for Optimal Data Analytics Performance," *International Journal of Computer Sciences and Engineering*, vol. 11, no. 10, pp. 51-58, 2023.
- [20] **O. Alzaabi, K. Mahri, M. El Khatib, and N. Alkindi**, "How Big Data Analytics Supports Project Manager in Project Risk Management – Cases from UAE Health Sector," *International Journal of Business Analytics and Security (IJBAS)*, vol. 3, no. 1, pp. 11-26, 2023, doi: 10.54489/ijbas.v3i1.201.
- [21] **G. Gkioka, T. Bothos, B. Magoutas, and G. Mentzas**, "Data Analytics Methods to Measure Service Quality: A Systematic Review," *Intelligent Decision Technologies*, vol. 17, no. 2, pp. 1-23, 2023, doi: 10.3233/IDT-230363.
- [22] **S. R. Janani, R. Subramanian, S. Karthik, and C. Vimalarani**, "Healthcare Monitoring Using Machine Learning Based Data Analytics," *International Journal of Computers Communications & Control*, vol. 18, no.1, 2023, doi: 10.15837/ijccc.2023.1.4973.
- [23] **S. H. Lee, H. Lee, and J. H. Kim**, "Enhancing the Prediction of User Satisfaction with Metaverse Service Through Machine Learning," *Computers*,

-
- Materials & Continua*, vol. 72, no. 3, pp. 4983-4997, 2022, doi: 10.32604/cmc.2022.027943.
- [24] **F. Esposito, A. Mikhailov, and N. Toussaint**, "Role of Data Analytics in Supply Chain for Improving Customer Satisfaction and Profitability," *Business & IT*, vol. 12, no. 1, pp. 108-115, 2022, doi: 10.14311/bit.2022.01.13.
- [25] **M. Helae, D. Ebrahimi, and F. Alzhouri**, "Data Analytics in the Pharmacology Domain," *International Journal of Big Data and Analytics in Healthcare*, vol. 7, no. 1, pp. 1-16, 2022, doi: 10.4018/ijbdah.314229.
- [26] **K. Batko and A. Ślęzak**, "The Use of Big Data Analytics in Healthcare," *Journal of Big Data*, vol. 9, no. 1, pp.1-24. 2022, doi: 10.1186/s40537-021-00553-4.
- [27] **A. Ikegwu, H. Nweke, C. Anikwe, U. Alo, and O. Okonkwo**, "Big Data Analytics for Data-Driven Industry: A Review of Data Sources, Tools, Challenges, Solutions, and Research Directions," *Cluster Computing*, vol. 25, no. 2, 2022, doi: 10.1007/s10586-022-03568-5.
- [28] **N. Verma, S. Singh, and D. Prasad**, "Machine Learning and IoT-Based Model for Patient Monitoring and Early Prediction of Diabetes," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 24, pp. 1-17. 2022, doi: 10.1002/cpe.7219.
- [29] **Aattouchi, S. el Mendili, and F. Elmendili**, "Sentiment Analysis of Health Care: Review," *E3S Web of Conferences*, vol. 319, no. 1, pp. 1-7, 2021.
- [30] **O. Salman, Z. Taha, M. Alsabah, Y. Hussein, A. Mohammed, and M. Aal-nouman**, "A Review On Utilizing Machine Learning Technology in The Fields of Electronic Emergency Triage and Patient Priority Systems in Telemedicine: Coherent Taxonomy, Motivations, Open Research Challenges and Recommendations for Intelligent Future Work," *Computer Methods and Programs in Biomedicine*, vol. 209, no.9, pp. 106357, 2021, doi: 10.1016/j.cmpb.2021.106357.
-

-
- [31] **H. Habehh and S. Gohel**, "Machine Learning in Healthcare," *Current Genomics*, vol. 22, no. 4, pp. 291-300, 2021, doi: 10.2174/1389202922666210705124359.
- [32] **Ahmed, M. Ahmad, G. Jeon, and F. Piccialli**, "A Framework for Pandemic Prediction Using Big Data Analytics," *Big Data Research*, vol. 25, no. 4, pp. 1-8, 2021, doi: 10.1016/j.bdr.2021.100190.
- [33] **M. Saini and J. Singh**, "Big Data Analytics and Machine Learning: Personalized, Predictive Health and Boost Exactitude Medicine Research," *International Journal of Management, IT & Engineering*, vol. 11, no. 12, pp. 47-56, 2021.
- [34] **D. U. Devesh, D. Kapoor, D. Patel, and Pharmacy**, "Repercussions of Big Data Analytics in Healthcare - An Updated Review," vol. 3, no. 2, pp. 2581-8880, 2021.
- [35] **V. Vargas-Calderón, M. Moros Ochoa, G. Castro Nieto, and J. Camargo**, "Machine Learning for Assessing Quality of Service in the Hospitality Sector Based on Customer Reviews," *Information Technology & Tourism*, vol. 23, no. 4, pp. 1-12, 2021.
- [36] **M. Hosseinzadeh, E. Azhir, O. Ahmed, M. Ghafour, S. Ahmed, A. Rahmani, and B. Vo**, "Data Cleansing Mechanisms and Approaches for Big Data Analytics: A Systematic Study," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 4, pp. 1-13, 2021, doi: 10.1007/s12652-021-03590-2.
- [37] **H. J. Lee, M. Lee, H. Lee, and R. Cruz**, "Mining Service Quality Feedback from Social Media: A Computational Analytics Method," *Government Information Quarterly*, vol. 38, no. 3, pp. 101571, 2021.
- [38] **A. I. Rahim, M. I. Ibrahim, K. I. Musa, S. L. Chua, and N. M. bin Yaacob**, "Assessing Patient-Perceived Hospital Service Quality and Sentiment in Malaysian Public Hospitals Using Machine Learning and Facebook
-

- Reviews," *International Journal of Environmental Research and Public Health*, vol. 18, no. 18, pp. 9912, 2021, doi: 10.3390/ijerph18189912.
- [39] **Nawab, G. Ramsey, and R. Schreiber**, "Natural Language Processing to Extract Meaningful Information from Patient Experience Feedback," *Applied Clinical Informatics*, vol. 11, no. 02, pp. 242-252, 2020, doi: 10.1055/s-0040-1708049.
- [40] **Shah, X. Yan, A. Shah, and G. Mamirkulova**, "Mining Patient Opinion to Evaluate the Service Quality in Healthcare: A Deep-Learning Approach," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 7, pp. 1-18, 2020, doi: 10.1007/s12652-019-01434-8.
- [41] **N. Zaman, D. Goldberg, A. Abrahams, and R. Essig**, "Facebook Hospital Reviews: Automated Service Quality Detection and Relationships with Patient Satisfaction," *Decision Sciences*, vol. 52, no. 6, pp. 1-25, 2020, doi: 10.1111/deci.12479.
- [42] **W. Wassouf, K. Salloum, and S. Balloul**, "Predictive Analytics Using Big Data for Increased Customer Loyalty: Syriatel Telecom Company Case Study," *Journal of Big Data*, vol. 7, no. 1, pp. 1-8. 2020, doi: 10.1186/s40537-020-00290-0.
- [43] **Kasten**, "Big Data Applications in Healthcare Administration," *International Journal of Big Data and Analytics in Healthcare*, vol. 5, no. 2, pp. 12-37, 2020, doi: 10.4018/IJBDAH.2020070102.
- [44] **F. Lopez, E. Núñez Valdez, V. García Díaz, and Z. Bursac**, "A Case Study for a Big Data and Machine Learning Platform to Improve Medical Decision Support in Population Health Management," *Algorithms*, vol. 13, no. 4, pp. 102, 2020, doi: 10.3390/a13040102.
- [45] **A. Zunic, P. Corcoran, and I. Spasic**, "Sentiment Analysis in Health and Wellbeing: A Systematic Review (Preprint)," *JMIR Medical Informatics*, vol. 8, no. 1, pp. 1-22, 2019, doi: 10.2196/16023.
-

- [46] **G. Thangarasu and K. Subramanian**, "Big Data Analytics for Improved Care Delivery in the Healthcare Industry," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 15, no. 10, p. 40, 2019, doi: 10.3991/ijoe.v15i10.10875.
- [47] **S. Abidi and S. Abidi**, "Intelligent Health Data Analytics: A Convergence of Artificial Intelligence and Big Data," *Healthcare Management Forum*, vol. 32, no.3, pp. 084047041984613, 2019, doi: 10.1177/0840470419846134.
- [48] **S. Kanimozhi, V. Nishok, D. Pavithra, and R. Pavithra**, "Opinion Mining to Analyze Drug Satisfaction of Patients," *International Journal for Scientific Research & Development*, vol. 6, no. 1, pp. 1266-1269, 2018.
- [49] **G. Koren, G. Nordon, K. Radinsky, and V. Shalev**, "Machine Learning of Big Data in Gaining Insight into Successful Treatment of Hypertension," *Pharmacology Research & Perspectives*, vol. 6, no. 3, pp. 1-6, 2018, doi: 10.1002/prp2.396.
- [50] **Wang and C. Alexander**, "Machine Learning in Big Data," *International Journal of Mathematical, Engineering and Management Sciences*, vol. 1, no. 2, pp. 52-61, 2016, doi: 10.33889/IJMEMS.2016.1.2-006.
- [51] **Archenaa and M. Anita**, "A Survey of Big Data Analytics in Healthcare and Government," *Procedia Computer Science*, vol. 50, no.2, pp. 408-413, 2015, doi: 10.1016/j.procs.2015.04.021.
- [52] **Na and W. Kyaing**, "Sentiment Analysis of User-Generated Content on Drug Review Websites," *Journal of Information Science Theory and Practice*, vol. 3, no. 1, pp. 6-23, 2015, doi: 10.1633/JISTaP.2015.3.1.1.
- [53] **J. Ebenezer and D. Sivan**, "Big Data Analytics in Healthcare: A Survey," *ARPN Journal of Engineering and Applied Sciences*, vol. 10, no. 2, pp. 3645-3650, 2015.
- [54] **S. Tan, G. Gao, and S. Koch**, "Big Data and Analytics in Healthcare," *Methods of Information in Medicine*, vol. 54, no. 6, pp. 546-547, 2015, doi: 10.3414/ME15-06-1001.

- [55] **A. Hopper and M. Uriyo**, "Using Sentiment Analysis to Review Patient Satisfaction Data Located on the Internet," *Journal of Health Organization and Management*, vol. 29, no. 2, pp. 221-233, 2015, doi: 10.1108/JHOM-12-2011-0129.
- [56] **S. R. Sukumar, R. Natarajan, and R. Ferrell**, "Quality of Big Data in Health Care," *International Journal of Health Care Quality Assurance*, vol. 28, no. 6, pp. 621-634, 2015.
- [57] **V. Dhar**, "Big Data and Predictive Analytics in Health Care," *Big Data*, vol. 2, no. 3, pp. 113-116, 2014, doi: 10.1089/big.2014.1525.
- [58] **F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi, and L. Donaldson**, "Use of Sentiment Analysis for Capturing Patient Experience From Free-Text Comments Posted Online," *Journal of Medical Internet Research*, vol. 15, no. 11, pP. 239, 2013, doi: 10.2196/jmir.2721.
- [59] **Raposo, H. Alves, and P. Duarte**, "Dimensions of Service Quality and Satisfaction in Healthcare: A Patient's Satisfaction Index," *Service Business*, vol. 3, no. 1, pp. 85-100, 2013, doi: 10.1007/s11628-008-0055-1.
- [60] **F. Alemi, M. Torii, L. Clementz, and D. Aron**, "Feasibility of Real-Time Satisfaction Surveys Through Automated Analysis of Patients' Unstructured Comments and Sentiments," *Quality Management in Health Care*, vol. 21, no. 1, pp. 9-19, 2012, doi: 10.1097/QMH.0b013e3182417fc4.
- [61] **J. Pollettini, S. Panico, J. Daneluzzi, R. Tinós, J. Baranauskas, and A. Macedo**, "Using Machine Learning Classifiers to Assist Healthcare-Related Decisions: Classification of Electronic Patient Records," *Journal of Medical Systems*, vol. 36, no. 6, pp. 3861-3874, 2012, doi: 10.1007/s10916-012-9859-6.

BIBLIOGRAPHY

- [1] **T. V. Iyelolu, E. E. Agu, C. Idemudia, and T. I. Ijomah**, “Leveraging artificial intelligence for personalized marketing campaigns to improve conversion rates,” *International Journal of Engineering Research and Development*, vol. 20, no. 8, pp. 253–270, 2024.
- [2] **M. R. Ayyagari**, “A framework for analytical CRM assessments challenges and recommendations,” *International Journal of Information, Business and Management*, vol. 13, no. 2, pp. 108–121, 2021.
- [3] **K. W. Walker**, “Exploring adaptive boosting (AdaBoost) as a platform for the predictive modeling of tangible collection usage,” *The Journal of Academic Librarianship*, vol. 47, no. 6, p. 102450, 2021.
- [4] **Sumit**, “Drug review sentiment analysis using boosting algorithms,” *International Journal of Trend in Scientific Research and Development (IJTSRD)*, vol. 5, no. 4, pp. 937–941, 2021.
- [5] **V. Djurisic, L. Kascelan, and S. Rogic**, “Bank CRM optimization using predictive classification based on the support vector machine method,” *Applied Artificial Intelligence*, vol. 34, no. 1, pp. 1–22, 2020.
- [6] **B. Zorić**, “Benefits of educational data mining,” *Journal of International Business Research and Marketing*, vol. 6, no. 1, pp. 12–16, 2020.
- [7] **Y. Rohani, Z. Torabi, and S. Kianian**, “A novel hybrid genetic algorithm to predict students' academic performance,” *Journal of Electrical and Computer Engineering Innovations*, vol. 8, no. 2, pp. 219–232, 2020.
- [8] **F. Ofori, E. Maina, and R. Gitonga**, “Using machine learning algorithms to predict students' performance and improve learning outcomes: A literature-based review,” *Journal of Information and Technology*, vol. 4, no. 1, pp. 33–55, 2020.

- [9] **D. Buenaño-Fernández, D. Gil, and S. Luján-Mora**, “Application of machine learning in predicting performance for computer engineering students: A case study,” *Sustainability*, vol. 11, no. 10, pp. 2833, 2019.
- [10] **R. P. Bunker and F. Thabtah**, “A machine learning framework for sport result prediction,” *Applied Computing and Informatics*, vol. 15, no. 1, pp. 27–33, 2019.
- [11] **N. Chen, C. Xiong, W. Du, C. Wang, X. Lin, and Z. Chen**, “An improved genetic algorithm coupling a back-propagation neural network model (IGA-BPNN) for water-level predictions,” *Water*, vol. 11, no. 9, p. 1795, 2019.
- [12] **Sahoo et al.**, “Deep learning-based health recommender system using collaborative filtering,” *Computation*, vol. 7, no. 2, pp. 1283–1299, 2019.
- [13] **V. Deshmukh et al.**, “Comparative study of different approaches of opinion mining,” *International Journal of Research*, vol. 8, no. 1, pp. 644–651, 2019.
- [14] **M. Abdullah and A. Fahad**, “Enhancing performance of educational data using big data and Hadoop,” *International Journal of Applied Engineering Research*, vol. 14, no. 19, pp. 3814–3819, 2019.
- [15] **Al-Homery, H. Asharai, and A. Ahmad**, “The core components and types of CRM,” *Pakistan Journal of Humanities and Social Sciences*, vol. 7, no. 1, pp. 121–145, 2019.
- [16] **E. T. Lau, L. Sun, and Q. Yang**, “Modeling, prediction, and classification of student academic performance using artificial neural networks,” *SN Applied Sciences*, vol. 1, no. 9, pp. 1–10, 2019.
- [17] **L. M. A. Zohair**, “Prediction of student’s performance by modeling small dataset size,” *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, pp. 1–18, 2019.
- [18] **B. K. Francis and S. S. Babu**, “Predicting academic performance of students using a hybrid data mining approach,” *Journal of Medical Systems*, vol. 43, no. 6, pp. 1–15, 2019.

- [19] **E. B. Belachew and F. A. Gobena**, “Student performance prediction model using machine learning approach: The case of Wolkite University,” *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 7, no. 2, pp. 46–50, 2017.
- [20] **V. Singh et al.**, “Opinion mining and analysis of movie reviews,” *Indian Journal of Science and Technology*, vol. 10, no. 19, pp. 10–14, 2017.
- [21] **V. Gopalakrishnan et al.**, “Patient opinion mining to analyze drug satisfaction using supervised learning,” *Journal of Applied Research and Technology*, vol. 15, no. 4, pp. 311–319, 2017.
- [22] **R. K. Rathore and J. Jayanthi**, “Student prediction system for placement training using fuzzy inference system,” *ICTACT Journal on Soft Computing*, vol. 7, no. 4, pp. 1473–1446, 2017.
- [23] **S. Agrawal, S. K. Vishwakarma, and A. K. Sharma**, “Using data mining classifier for predicting student’s performance in UG level,” *International Journal of Computer Applications*, vol. 172, no. 8, pp. 39–44, 2017.
- [24] **S. A. Oloruntoba and J. L. Akinode**, “Student academic performance prediction using support vector machine,” *International Journal of Engineering Sciences and Research Technology*, vol. 6, no. 12, pp. 588–597, 2017.
- [25] **N. Arora and N. Singh**, “Factors affecting the academic performance of college students,” *Journal of Educational Technology*, vol. 14, no. 1, pp. 47–53, 2017.
- [26] **B. Arputhamary et al.**, “Interpreting the public sentiment variations on Twitter using virtual shuffling for data movement in big data,” *International Journal of Control Theory and Applications*, vol. 9, no. 26, pp. 389–395, 2016.
- [27] **G. Biau et al.**, “A random forest guided tour,” *Test*, vol. 25, no. 2, pp. 197–227, 2016.

- [28] **B. M. Bandgar et al.**, “Analysis of real-time social tweets for opinion mining,” *International Journal of Applied Engineering Research*, vol. 11, no. 2, pp. 1404–1407, 2016.
- [29] **I. R. Jayasekara et al.**, “Opinion mining of customer reviews: Feature and smiley based approach,” *International Journal of Data Mining & Knowledge Management Process*, vol. 6, no. 1, pp. 01–11, 2016.
- [30] **M. Aarati Mahadik et al.**, “Aspect-based opinion mining for identifying customer preferences,” *International Journal of Engineering Research and Applications*, vol. 6, no. 2, 2016.
- [31] **J. Vaiz and M. Ramaswami**, “A study on technical indicators in stock price movement prediction using decision tree algorithms,” *American Journal of Engineering Research (AJER)*, vol. 5, no. 12, pp. 207–212, 2016.
- [32] **L. Wang**, “Data mining, machine learning and big data analytics,” *International Transaction of Electrical and Computer Engineers System*, vol. 4, no. 2, pp. 55–61, 2017.
- [33] **M. Rawat et al.**, “A new alley in opinion mining using Senti Audio Visual Algorithm,” *International Journal of Engineering Research and Applications*, vol. 6, no. 2, pp. 01–09, 2016.
- [34] **R. Sumitha, E. S. Vinothkumar, and P. Scholar**, “Prediction of students outcome using data mining techniques,” *International Journal of Scientific Engineering and Applied Science (IJSEAS)*, vol. 2, no. 6, 2016.
- [35] **P. Baranikumar et al.**, “Feature extraction of opinion mining using ontology,” *International Journal of Advanced Computer and Electronics Engineering*, vol. 1, no. 1, pp. 18–22, 2016.
- [36] **R. Rajput et al.**, “Review of sentimental analysis methods using lexicon-based approach,” *International Journal of Computer Science and Mobile Computing*, vol. 5, no. 2, pp. 159–166, 2016.

- [37] **A. H. Yazdavar, M. Ebrahimi, and N. Salim**, “Fuzzy-based implicit sentiment analysis on quantitative sentences,” *Journal of Soft Computing and Decision Support Systems*, vol. 3, no. 4, pp. 7–18, 2016.
- [38] **W. Zhang et al.**, “Predicting potential side effects of drugs by recommender methods and ensemble learning,” *Neurocomputing*, vol. 173, pp. 979–987, 2016.
- [39] **A. Algarni**, “Data mining in education,” *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 6, pp. 456–461, 2016.
- [40] **A. Dey**, “Machine Learning Algorithms: A Review,” *International Journal of Computer Science and Information Technologies*, vol. 7, no. 3, pp. 1174–1179, 2016.
- [41] **L. Haldurai, T. Madhubala, and R. Rajalakshmi**, “A study on genetic algorithm and its applications,” *International Journal of Computer Sciences and Engineering*, vol. 4, no. 10, p. 139, 2016.
- [42] **F. O. Isinkaye et al.**, “Recommendation systems: Principles, methods and evaluation,” *Egyptian Informatics Journal*, vol. 16, no. 3, pp. 261–273, 2015.
- [43] **J. Van der Zwart et al.**, “Pre-occupancy evaluation of patient satisfaction in hospitals,” *HERD: Health Environments Research & Design Journal*, vol. 9, no. 1, pp. 110–124, 2015.
- [44] **H. Iyer**, “Sentiment analysis for visuals using natural language processing,” *International Journal of Computer Applications*, vol. 128, no. 6, pp. 31–35, 2015.
- [45] **S. Mehta**, “Patient satisfaction reporting and its implications for patient care,” *The AMA Journal of Ethics*, vol. 17, no. 7, pp. 616–621, 2015.
- [46] **A. Price et al.**, “Examining the role of patient experience surveys in measuring health care quality,” *Medical Care Research and Review*, vol. 71, no. 5, pp. 522–554, 2014.

- [47] **Turdakov et al.**, “A framework for text analysis,” *Program. Computer Software*, vol. 40, no. 5, pp. 288–295, 2014.
- [48] **C. M. Wittich et al.**, “Medication errors: An overview for clinicians,” *Mayo Clinic Proceedings*, vol. 89, no. 8, pp. 1116–1125, 2014.
- [49] **W. U. et al.**, “Exploiting online discussions to discover unrecognized drug side effects,” *Methods of Information in Medicine*, vol. 52, no. 2, pp. 152–159, 2013.
- [50] **R. Priya, B. F. de Souza, A. L. Rossi, and A. C. de Carvalho**, “Predicting execution time of machine learning tasks for scheduling,” *International Journal of Hybrid Intelligent Systems*, vol. 10, no. 1, pp. 23–32, 2013.
- [51] **T. M. Lakshmi, A. Martin, and V. P. Venkatesan**, “An analysis of students’ performance using genetic algorithm,” *Journal of Computer Sciences and Applications*, vol. 1, no. 4, pp. 75–79, 2013.
- [52] **A. Zgierska et al.**, “Patient satisfaction, prescription drug abuse, and potential unintended consequences,” *JAMA*, vol. 307, no. 13, pp. 1377–1378, 2012.
- [53] **F. Alemi et al.**, “Feasibility of real-time satisfaction surveys through automated analysis of patients’ unstructured comments and sentiments,” *Quality Management in Health Care*, vol. 21, no. 1, pp. 9–19, 2012.
- [54] **M. S. Upadhya**, “Fuzzy logic-based evaluation of performance of students in colleges,” *Journal of Computer Applications (JCA)*, vol. 5, no. 1, 2012.
- [55] **M. Ramaswami and R. Rathinasabapathy**, “Student performance prediction,” *International Journal of Computational Intelligence and Informatics*, vol. 1, no. 4, pp. 231–235, 2012.
- [56] **M. S. Farooq, A. H. Chaudhry, M. Shafiq, and G. Berhanu**, “Factors affecting students’ quality of academic performance: A case of secondary school level,” *Journal of Quality and Technology Management*, vol. 7, no. 2, pp. 1–14, 2011.

- [57] **J. M. Hush**, “Patient satisfaction with musculoskeletal physical therapy care: A systematic review,” *Physiotherapy*, vol. 91, no. 1, pp. 25–36, 2011.
- [58] **Q. Zhang, L. Cheng, and R. Boutaba**, “Cloud computing: State-of-the-art and research challenges,” *Journal of Internet Services and Applications*, vol. 1, no. 1, pp. 7–18, 2010.
- [59] **P. Greenberg**, “The impact of CRM 2.0 on customer insight,” *Journal of Business & Industrial Marketing*, vol. 25, no. 6, pp. 410–419, 2010.
- [60] **Michael Kuhn et al.**, “A Side Effect Resource to Capture Phenotypic Effects of Drugs” *Molecular Systems Biology* 6(1) 343. 2010.
- [61] **Jiang X, Yi Z, and Lv JC**, “Fuzzy SVM with a New Fuzzy Membership Function” *Neural Computing & Applications* 15(3) 268–276. 2006
- [62] **O'Toole RV et al.**, “LEAP Study Group. Determinants of Patient Satisfaction After Severe Lower-Extremity Injuries” *Journal of Bone Joint Surgery America* 90(6), 1206–1211. 2008.
- [63] **Yoo I and Song M**, “Biomedical Ontologies and Text Mining for Biomedicine and Healthcare: A Survey” *Journal of Computing Science and Engineering* 2(2) 109–136. 2008.
- [64] **Williams ML, Waldauer C, and Duggal VG**, “Gender Differences in Economic Knowledge: An Extension of the Analysis” *The Journal of Economic Education* 23(3) 219–231. 1992.
- [65] **Gustafsson A, Johnson MD, and Roos I**, “The Effects of Customer Satisfaction, Relationship Commitment Dimensions, and Triggers on Customer Retention” *Journal of Marketing* 69(4) 210–218. 2005.
- [66] **Beaumont-Walters Y and Soyibo K**, “An Analysis of High School Students' Performance on Five Integrated Science Process Skills” *Research in Science & Technological Education* 19(2) 133–145. 2001.

- [67] **Epstein KR**, “Patients’ Perceptions of Office Medical Practice: Judging Quality Through the Patients’ Eyes” *American Journal of Medical Quality* 11(2) 73–80. 1996.
- [68] **Koran LM et al.**, “Medical Evaluation of Psychiatric Patients. I. Results in a State Mental Health System” *Archives of General Psychiatry* 46(8) 733–740. 1989.

BOOK

- [1] **J. A. W. Joseph, J. S. Vaiz, and R. Muruges**, “Modeling Cognitive Load in Mobile Human-Computer Interaction Using Eye Tracking Metrics,” in *Advances in Artificial Intelligence, Software and Systems Engineering. AHFE 2021, Lecture Notes in Networks and Systems*, vol. 271, T. Z. Ahram, W. Karwowski, and J. Kalra, Eds. Cham: Springer, 2021.
- [2] **Y. Xiaogao and P. Ruiqing**, “Research on Big Data-Driven High-Risk Students Prediction,” in *2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, pp. 145–149. 2017.
- [3] **H. Duan, P. Li, Y. Shi, X. Zhang, and C. Sun**, “Interactive Learning Environment for Bio-Inspired Optimization Algorithms for UAV Path Planning,” *IEEE Trans. Educ.*, vol. 58, no. 4, pp. 276–281, 2015.
- [4] **T. Brummett, P. Sheinidashtegol, D. Sarkar, and M. Galloway**, “Performance Metrics of Local Cloud Computing Architectures,” in *2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing*, pp. 25–30. 2015
- [5] **R. Kune, P. K. Konugurthi, A. Agarwal, R. R. Chillarige, and R. Buyya**, “Genetic Algorithm-Based Data-Aware Group Scheduling for Big Data Clouds,” in *2014 IEEE/ACM International Symposium on Big Data Computing*, pp. 96–104. 2014

- [6] **V. Miškovic**, “Machine Learning of Hybrid Classification Models for Decision Support,” in *Sinteza 2014 - Impact of the Internet on Business Activities in Serbia and Worldwide*, pp. 318–323. 2014,
- [7] **R. S. Yadav, A. K. Soni, and S. Pal**, “A Study of Academic Performance Evaluation Using Fuzzy Logic Techniques,” in *2014 International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 48–53. 2014
- [8] **K. Li, C. Gibson, D. Ho, Q. Zhou, J. Kim, O. Buhisi, and M. Gerber**, “Assessment of Machine Learning Algorithms in Cloud Computing Frameworks,” in *2013 IEEE Systems and Information Engineering Design Symposium*, pp. 98–103. 2013
- [9] **S. A. Ajila and A. A. Bankole**, “Cloud Client Prediction Models Using Machine Learning Techniques,” in *2013 IEEE 37th Annual Computer Software and Applications Conference*, pp. 134–142. 2013
- [10] **J. Kennedy and R. C. Eberhart**, “Particle Swarm Optimization,” in *Proc. IEEE Int. Conf. Neural Networks*, pp. 1942–1948. 1995
- [11] **R. J. Vidmar**, “On the Use of Atmospheric Plasmas as Electromagnetic Reflectors,” *IEEE Trans. Plasma Sci.*, vol. 21, no. 3, pp. 876–880, 1992.
- [12] **G. W. Juetten and L. E. Zeffanella**, “Radio Noise Currents in Short Sections on Bundle Conductors,” presented at the *IEEE Summer Power Meeting*, Dallas, TX, 1990.



IAIR CONFERENCES

National Conference on Latest Innovations in Engineering, Science, Management and Humanities (NCLIESMH- 2024)

CERTIFICATE

NCLIESMH/2024/C0524572

This is to certify that Nayan Sarkar, Research Scholar,

Department of Computer Science and Engineering, P.K.University, Shivpuri,

M.P., India.

**presented a paper in the National Conference on Latest Innovations in Engineering, Science
Management and Humanities (NCLIESMH- 2024)**

at The Downton Hotel, Raipur, Chhattisgarh, India on 26th May 2024

Title of the Paper Navigating The Future of Sentiment Analysis in Healthcare : From

Lexicons to Real-Time Insights.

K. Jayakumar.

Dr. K. Jaya Kumar

Session Chair



A. Usha Ruby

Dr. A. Usha Ruby

Session Chair



IARF CONFERENCES

International Conference on Researches in Engineering, Science Management and Humanities (ICRESMH- 2025)

ICRESMH/2025/IC04254/2

Certificate

<https://iarfconferences.com>

This is to certify that Nayan Sarkar,

Research scholar, Department of Computer science and Engineering,

P.K. University, Shivpur, M.P., India.

presented a paper in the International Conference on Researches in Engineering, Science,
Management and Humanities (ICRESMH – 2025)
at **Le Meridien Gurgaon, Delhi, India on 27th April 2025**

Title of the Paper Enhancing sentiment classification Accuracy in Medical Datasets
Using A soft voting Ensemble of supervised Machine Learning Algorithms.

Nayan Sarkar

Dr. Harold Andrew Patrick

Session Chair



Shabnam Arora

Dr. Shabnam Arora

Session Chair

A Study of Hybrid Sentiment Analysis of Patient Feedback using Lexicon-Based and Machine Learning Approaches

Nayan Sarkar¹, Dr. Balveer Singh²

^{1, 2} Department of Computer Science & Engineering, P.K. University, Shivpuri, M.P., India

Abstract

The keyword-based approach in sentiment analysis offers a structured and interpretable method to assess patient feedback, aiming to enhance patient satisfaction and healthcare quality. This research proposes a hybrid sentiment analysis framework for evaluating patient feedback in the healthcare sector. The study employs both keyword-based and machine-learning approaches to classify patient comments as positive, negative, or neutral. Data was collected from the “RateMDs” website, containing physician profiles and patient reviews. The preprocessing phase involved text normalization, removal of noise, punctuation, numbers, and stop words, followed by transformation into a structured dataset for analysis. The R-Tool was used for implementing sentiment analysis, supported by positive and negative lexicons derived from publicly available repositories. Machine learning techniques, including Bayesian classifiers and the Weka data mining tool, were applied for accurate classification. The system architecture integrates preprocessing, sentiment scoring, and result visualization. Results indicate that the proposed model effectively identifies patient sentiments, helping hospitals and physicians assess service quality, patient satisfaction, and areas for improvement.

Keywords: Sentiment Analysis, Patient Satisfaction, Healthcare Quality, Machine Learning, Patient Feedback.

Date of Submission: 25-10-2025

Date of Acceptance: 05-11-2025

I. Introduction

In the evolving landscape of healthcare, understanding patient emotions, opinions, and experiences has become a vital component in improving service quality and patient-centered care. Traditional feedback mechanisms, such as surveys and satisfaction questionnaires, provide valuable insights but often fail to capture the depth of patients’ emotions and sentiments embedded within their narratives. With the rise of digital transformation and the widespread use of online healthcare platforms, patients today share their experiences openly on hospital websites, review portals, and social media. This vast pool of textual data presents both an opportunity and a challenge for healthcare providers—to meaningfully interpret these sentiments and translate them into actionable insights. Here, sentiment analysis, a subfield of Natural Language Processing (NLP) and artificial intelligence (AI), emerges as a powerful analytical tool for understanding patient satisfaction and enhancing healthcare quality. Sentiment analysis refers to the computational study of opinions, emotions, and attitudes expressed in text. By employing machine learning and deep learning algorithms, sentiment analysis models classify text as positive, negative, or neutral, thereby revealing the underlying emotional tone of patient feedback. In the context of healthcare, this technology enables hospitals and policymakers to analyze patient reviews, detect dissatisfaction trends, and respond proactively. Unlike conventional survey data, sentiment analysis captures spontaneous and unfiltered expressions of patient experiences, offering a more authentic representation of satisfaction levels. This technology also bridges the communication gap between patients and healthcare providers, allowing for continuous monitoring of patient perceptions in real time.

The importance of patient satisfaction in healthcare cannot be overstated. It is widely recognized as a key indicator of healthcare quality, patient safety, and institutional performance. Satisfied patients are more likely to adhere to medical advice, maintain follow-up appointments, and recommend healthcare providers to others. Conversely, dissatisfaction can lead to mistrust, treatment noncompliance, and negative public perception. Therefore, understanding patient satisfaction is not merely a public relations exercise—it is a cornerstone of effective healthcare delivery and organizational success. Sentiment analysis enables institutions to systematically evaluate these satisfaction levels and uncover specific areas for improvement, such as waiting time, staff behavior, communication clarity, or treatment efficacy. Recent advancements in AI and machine learning have significantly improved the accuracy and applicability of sentiment analysis in healthcare. Algorithms such as Support Vector Machines (SVM), Naïve Bayes, Recurrent Neural Networks (RNN), and Transformer-based architectures like BERT (Bidirectional Encoder Representations from Transformers) are

capable of processing complex linguistic patterns and contextual meanings. These models can identify subtle emotional cues in patient feedback, such as anxiety, gratitude, or frustration, which traditional analysis methods might overlook. Moreover, integrating sentiment analysis with big data analytics allows healthcare organizations to evaluate thousands of patient reviews across multiple platforms efficiently. This leads to evidence-based decision-making and data-driven strategies for quality enhancement.

In addition to improving service quality, sentiment analysis plays a crucial role in public health monitoring and policy formulation. By aggregating sentiment data from various sources, policymakers can gauge community-level perceptions of healthcare services, identify regions with lower satisfaction levels, and allocate resources accordingly. During crises, such as pandemics or health emergencies, sentiment analysis can also serve as an early-warning mechanism by tracking public fear, misinformation, and trust in healthcare institutions. Furthermore, hospitals can use sentiment analysis to evaluate the success of new healthcare programs, measure the impact of interventions, and assess patient trust in technology-driven care models such as telemedicine or AI-assisted diagnosis. The integration of sentiment analysis into healthcare systems aligns with the broader shift toward value-based healthcare, where outcomes and patient experiences hold greater significance than mere procedural volumes. By leveraging sentiment analysis, healthcare institutions can move beyond numerical ratings to understand the emotional context behind patient opinions. This emotional intelligence empowers healthcare administrators and clinicians to adopt more empathetic communication styles, develop patient-friendly environments, and tailor medical services to individual needs. Additionally, predictive sentiment analytics can anticipate potential dissatisfaction and trigger timely interventions, thereby preventing negative experiences from escalating.

However, the implementation of sentiment analysis in healthcare is not without challenges. Issues such as data privacy, linguistic ambiguity, algorithmic bias, and the need for domain-specific training data must be addressed to ensure ethical and accurate outcomes. Despite these challenges, the benefits far outweigh the limitations. As healthcare becomes increasingly digitalized, sentiment analysis will continue to evolve, offering more sophisticated insights through multimodal analysis—combining text, voice, and facial expressions—to capture the full spectrum of patient sentiment. Sentiment analysis represents a paradigm shift in how healthcare organizations measure and enhance patient satisfaction. By combining AI-driven text analytics with human empathy and clinical insight, healthcare providers can transform patient feedback into actionable intelligence. This not only fosters patient trust and engagement but also contributes to the overall improvement of healthcare quality, efficiency, and transparency.

II. Review of Related Studies

He, Mu. (2024) with the rise of e-health systems, the problems with conventional healthcare systems, such as inaccessibility, high operating costs, and inefficiency, have been greatly alleviated. Improving patient care and system efficiency are two of the primary goals of this paper's examination of Natural Language Processing (NLP) and its pivotal position in electronic health record (EHR) systems. Utilizing sentiment analysis to glean emotional information from patient experiences, the research applies natural language processing methods to sift through and make sense of massive amounts of healthcare data. Patient satisfaction and the quality of care provided may be improved by using this data to gauge how patients feel about various treatments and services. By detecting possible problems based on patients' emotional reactions, sentiment analysis also plays an important role in mental health monitoring. The findings point to the fact that public health initiatives, mental health monitoring, and patient happiness may all be greatly improved with the use of NLP-driven sentiment analysis. Healthcare professionals may improve healthcare outcomes by using sentiment analysis to better identify and react to community needs. When it comes to improving decision support, managing unstructured data more effectively, and increasing the accuracy of data analysis, this study offers crucial insights for the future development of e-health systems.

Rizwan Rashid et al., (2023) Traditional healthcare systems were plagued by inefficiency, high operating costs, and restricted accessibility; however, these issues have been greatly alleviated with the rise of e-health systems. Natural Language Processing (NLP) plays an essential role in e-health systems, and this article delves into that function, specifically looking at sentiment analysis and how it may revolutionize patient care and system efficiency. The research uses natural language processing methods to sift through mountains of healthcare data in search of patterns, and it makes use of sentiment analysis to glean insights about patients' subjective experiences with treatment. Improving care quality and patient happiness is possible with the use of this data, which reveals how patients feel about various treatments and services. Sentiment analysis is also vital in mental health monitoring since it may spot problems depending on how individuals are expressing themselves emotionally. Patient happiness, mental health monitoring, and public health strategy effectiveness may all be greatly improved using NLP-driven sentiment analysis, according to the findings. Better healthcare outcomes may be achieved when healthcare practitioners use sentiment analysis to get a better understanding of community needs and react accordingly. Particularly in the areas of unstructured data handling, data analysis

accuracy, and decision support, this study offers significant insights for the future development of e-health systems.

Ramyasri, V et al., (2019) The exponential growth of online communication platforms has given individuals all over the globe a platform to voice their thoughts and feelings on any given subject, service, or product. Patients are the ultimate arbiters of a healthcare service's success or failure. If most people who use the service are satisfied, then it's a success; otherwise, it requires some tweaks. It is necessary to examine consumer feedback in order to enhance the service. Extracting and evaluating online material by hand is a time-consuming and laborious process. As a result, the field of Sentiment Analysis was born. This practice is sometimes called opinion mining. Many health organizations are use it to make informed judgments about their service. In this work, we provide sentiment analysis as a tool for better healthcare by analyzing people' perceptions about hospitals. The implementation of this involves analyzing the sentiment using an approach based on a vocabulary.

Hopper, Anthony & Uriyo, Maria. (2015) This research aims to quantify text-based information found on the internet using sentiment analysis and time-to-next-complaint approaches. Equally important, the authors show how managers may arrange sentiment analysis data into meaningful information using time-to-next-complaint approaches; this information can then be shared with physicians and other personnel. The writers analyzed patient comments for a subset of Virginia gynecologists using sentiment analysis. In order to make sense of this data, the writers used time-to-next-complaint procedures among other ways. If healthcare administrators are interested in turning web-based material into meaningful, measurable information, the authors showed that sentiment analysis and time-to-next-complaint approaches could be helpful tools. The research isn't without its flaws. One problem is that gender, wealth, culture, and other socio-demographic biases in selection won't be taken into consideration by either the data collection or the methods used to evaluate it. The writers also failed to provide crucial information on the number of patients seen by the specified doctors. Lastly, healthcare administrators may be hesitant to use online data due to the difficulty in persuading clinicians to accept online remarks as actual. Using sentiment analysis and time-to-next-complaint approaches, the research shows how hospital managers may mine online patient comments for useful information. Clinic and doctor's office administrators may use sentiment analysis and time-to-next-complaint approaches to sift through online patient comments; this research is among the first to do so.

Khan, Taimoor & Khalid, Shehzad. (2015). In health care sentiment analysis, patients' self-reported issues with their treatment are the focus of the diagnostic process. By considering patient feedback, policies and changes may be developed to tackle patients' issues head-on. Sentiment analysis has matured and found new uses outside its original commercial product domain. In addition to making treatment and service recommendations, aspect based analysis of health care highlights the benefits of each option. An efficient and accurate judgment is reached by analyzing millions of review papers using machine learning algorithms. While supervised methods are quite accurate, they cannot be applied to domains that are unknown, while unsupervised methods are inaccurate. Since unsupervised methods are more applicable in this era of information deluge, there is a greater emphasis on improving their accuracy.

Greaves, Felix et al., (2013) Blogs, social networks, and physician rating websites all include vast volumes of unstructured, free-text material on healthcare quality that is not systematically collected. Sentiment analysis and other new forms of data analysis could help us make better use of this data to enhance healthcare quality. We looked at using machine learning to decipher patients' unstructured feedback on their treatment. We classified patients' online free-text remarks on their healthcare as good or unfavorable using sentiment analysis methods. After comparing the patient's free-text description with their quantitative evaluation of care, we attempted to automatically predict whether the hospital was clean, if they were treated with dignity, and if they would suggest the facility. In 2010, we used the data-mining tool Weka to analyze 6412 comments posted on the English National Health Service website on hospitals. We then used machine learning methods to these remarks. In addition, we used Spearman rank correlation to compare the sentiment analysis findings with the hospital-level results of the 161 acute adult hospital trusts in England's paper-based national inpatient survey. Comparing quantitative care ratings with those obtained from free-text comments using sentiment analysis, we find that 81% of patients felt the hospital was clean, 84% felt they were treated with dignity, and 89% would recommend the facility to others.

III. Proposed Methodology

Research Approach

One potential approach to this study is to collect patient feedback by hand. Users' comments, whether pleasant or bad. Following that, it will use two methods: a keyword-based approach and a machine-based technique.

To find representative qualities and distinguishing features, this case makes advantage of the classifier's load. A class's "representative attributes" are its outward manifestations, while its "distinctive characteristics"

are its internal details. With these categorization weights, the Bayesian algorithm can determine the likelihood and provide superior results. This two-stage mechanical sentiment analysis approach was developed by Barbosa et al. for the purpose of tweet classification. To reduce the labeling trial, they use a noisy build upset while creating classifiers. The first thing they do is sort tweets by method, whether subjective or objective. After that, we classify subjective tweets as either favorable or negative.

Based on normalized noisy tweet pronunciation, researchers have created a clustering approach. Words with the same utterance are grouped together using common tokens in an utterance-based word cluster. Tokens representing numbers, HTML links, user IDs, and the names of objective organizations are some of the text processing methods used for normalization. Following normalization, polarity lexicons were recognized using probabilistic models. The BoosTexter classifier achieved a lower error rate when trained with these polarity lexicons as features. A practical probability model for analyzing sentiment on Twitter was introduced by Wu et al. An action is valued and placed into helpful likelihood if they commence the usage of @tweetusername in the body of a tweet. Retweets that begin with @tweetusername generate motivated action and commit themselves to motivated likelihood. A robust relationship between these probabilities was discovered. Using the Twitter API and Emoticons, Pak et al. automated the process of collecting tweets and annotating them to create a Twitter corpus.

A method of machine learning that makes use of the data mining tool Weka to train an algorithm to categorize comments into distinct groups based on their similarities and differences. Weka, which has been employed in most prior studies, provides accurate categorization results and is also useful in healthcare.

Compare results to quantitative ratings obtained from similar individual patients on a Likert scale to validate accuracy. The next step is to analyze the patient's feedback using a three-question-and-answer format, such as "what I liked," "what could have been improved," and "any other comments." The patient will then be asked to rate the cleanliness of the hospital and the level of respect and dignity they received during their stay.

Sources of data

Data comes from a variety of places; for example, surveys, social media, healthcare websites, blogs, etc. We gathered our database from <https://www.ratemds.com> for this project. You may find extensive biographical information on doctors here, along with links to their areas of expertise and feedback from actual patients. The information on the doctor and hospital that is included in these remarks is feedback. There are a lot of doctors' profiles on our site, and they come from all around the globe.

The website "ratemds" also featured the specializations of Indian physicians from various cities in India. Select the specialization of the doctor, the city, and then either male or female to narrow your search for a doctor on the "ratemds" website. After that, we may see the profiles of doctors and read comments from their patients right here. More than sixty patient comments on physician feedback texts were culled from this website. The text data is compiled according to the medical specialties of the doctors. We will transfer this data from an Excel file to a "CSV" format so it may be processed later.

Pre-processing

Data pre-processing is a process in which processing is performed on raw data to organize it to remove noise or unwanted text for the purpose of data cleaning and for better classification. Pre-processing is like to removing unwanted text, punctuation, NA values, numbers, etc. Pre-processing begins with the Comments Review Database, which serves as the initial source of input data. The next stage is Preprocessing, where textual data undergoes several cleaning operations to prepare it for analysis. These preprocessing steps include converting all text to lowercase, removing stop words, eliminating punctuation marks, deleting numerical values, discarding missing (NA) entries, and removing hyperlinks. Once the data is cleaned, it moves to the Transformation phase, where the text is converted into a structured format suitable for computational analysis. After transformation, Feature Selection is performed to identify and extract the most relevant attributes or variables that significantly contribute to model accuracy and efficiency. Finally, the refined data is subjected to Classification, where machine learning or statistical algorithms categorize the processed comments into predefined classes or sentiment categories, completing the analytical workflow.

Pre-processing

In the pre-processing part, it was indicated that you should lowercase all content, remove punctuation marks, stop words, NA values, numerals, links, etc. This section discusses the infrastructure in detail that will be used to forecast healthcare outcomes based on patient input, namely online text comments.

System Architecture

In this paper, proposed architecture consists- R data analysis tool, sentiment analysis, and pre-processing workflow.

- a) **R-Tool:** R-Programming makes use of this tool. It is possible to create a software that can predict the emotion score of text using this tool.
- b) **Sentiment Analysis:** Here we show how to use the R programming language to determine the sentiment score of a patient's response.
- c) **Positive text:** Here you may find a library of positive words together with their synonyms.
- d) **Negative Text:** This text file contains both positive and negative words, as well as their synonyms.

IV. Results and Discussion

Sentiment Score

The emotion score must be counted after pre-processing. The R-tool is used to tally sentiment scores. The next step is to write an R program that will compare each comment's sentiment score to two text files, one containing positive comments and the other containing negative ones. This text file contains both positive and negative sentiments extracted from the "Twitter sentiment analysis tutorial" repository on Github. The sentiment score is obtained by comparing the text files containing positive and negative comments made by patients with the physician's database. After tallying all of the good and negative words in a statement, this score will display the emotion score. The sentiment score of a comment sample predicted using R programming in R-Tool. See that a score of 0 indicates no opinion, a score of -1 indicates a bad remark, and a score of 3 indicates a good comment.

Table 1 shows the total comments and the extracted comments organized in a categorized way. Find the doctor's number, the total number of comments, the percentage of positive remarks, the percentage of negative comments, and the percentage of neutral comments in this table. These remarks provide the patient's perspective on their doctor, which may be categorized as good, negative, or neutral.

Table 1: Classified commented feedback of patients

Sr. No.	Number of Comments	Negative Comments Count	Positive Comments Count	Neutral Comments Count
D1	25	7	15	3
D2	22	3	17	2
D3	21	5	13	3
D4	19	2	15	2
D5	15	1	12	2
D6	13	0	12	1
D7	11	2	8	1
D8	10	1	7	2

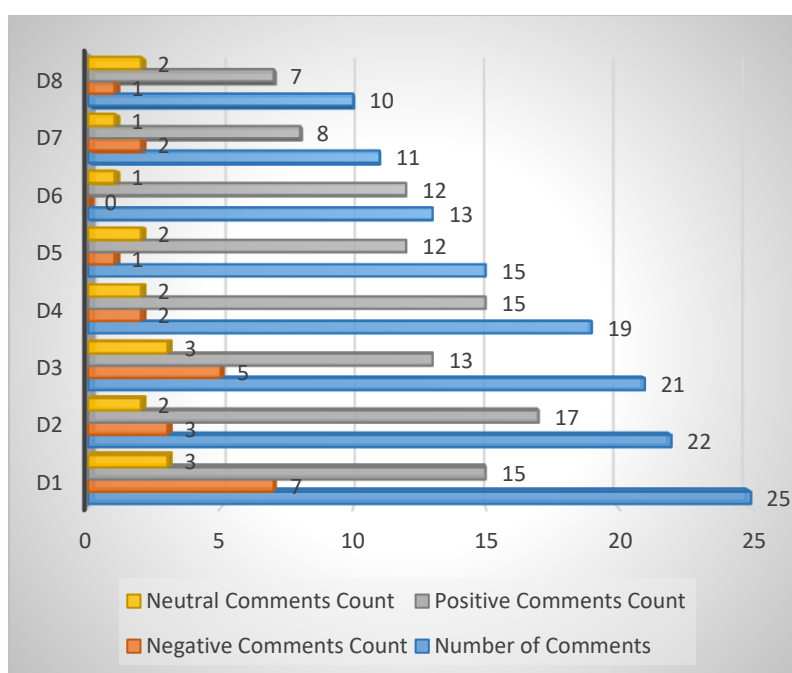


Figure 1: Distribution of Positive, Negative, and Neutral Comments Across Datasets (D1–D8) Showing Overall Comment Volume

Table 1 is shown graphically in Figure 1. The data shows that patients are more likely to have a good impression of their doctor when compared to those who are indifferent or have a bad impression.

Table 2: Percentage Distribution of Comment Sentiments Across Datasets

Dataset ID	% Negative Comments	% Positive Comments	% Neutral Comments
D1	28.0	64.0	8.0
D2	10.0	85.0	5.0
D3	22.0	65.0	13.0
D4	14.0	76.0	10.0
D5	5.0	80.0	15.0
D6	0.0	90.0	10.0
D7	20.0	78.0	2.0
D8	12.0	78.0	10.0

The table shows the fraction of patients' views based on the data. The percentage breakdown of comments on a doctor is shown in the table. The data in the table show that most patients have a good opinion of their doctor, with just a tiny minority holding a neutral or negative view. Unfortunately, some patients did not provide their good comments.

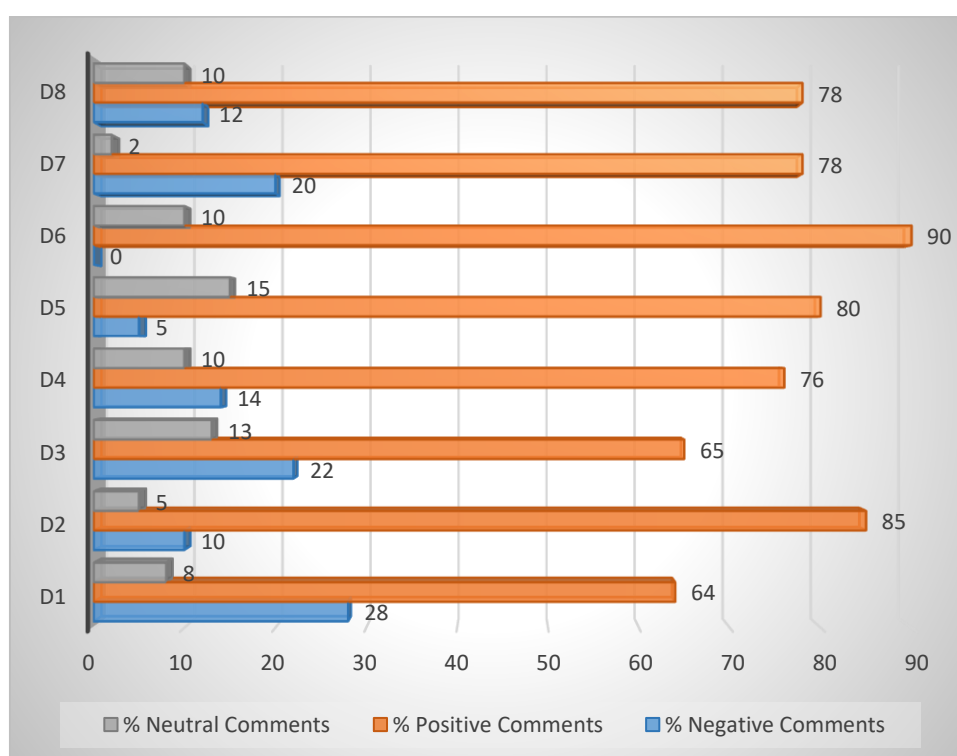


Figure 2: Distribution of Positive, Negative, and Neutral Comments Across Datasets (D1–D8) Showing Overall Comment Volume based on patients' opinions

Table 2's visual depiction is shown above Figure 2. Compared to negative and neutral graphs, positive graphs on patients' graphs had a greater rate of responses. Compared to the positive graph, the negative and neutral graphs make up a negligible amount.

V. Conclusion

In sentiment analysis serves as a transformative tool in the pursuit of patient-centered healthcare and continuous quality improvement. By analyzing patient feedback through advanced AI and machine learning algorithms, healthcare providers gain deeper insights into the emotional and experiential dimensions of care. This approach transcends traditional survey-based assessments, offering a more nuanced and authentic understanding of patient satisfaction. The ability to classify sentiments as positive, negative, or neutral enables healthcare institutions to identify service strengths, address areas of dissatisfaction, and implement targeted improvements. Furthermore, the integration of sentiment analysis into healthcare management promotes transparency, responsiveness, and data-driven decision-making. Although challenges related to data ethics, linguistic complexity, and algorithmic accuracy remain, ongoing advancements in NLP and AI continue to

refine these technologies. Ultimately, sentiment analysis empowers healthcare organizations to not only monitor satisfaction but also cultivate a culture of empathy, trust, and excellence—thereby enhancing the overall quality of healthcare delivery and strengthening the relationship between patients and providers.

References:

- [1] M. He, “Enhancing E-Health with natural language processing: The role of sentiment analysis in modern healthcare,” *Transactions on Computer Science and Intelligent Systems Research*, vol. 7, no. 1, pp. 285–290, 2024.
- [2] M. R. Rashid, A. Nawaz, T. Ali, and G. Mustafa, “Enhancing healthcare services recommendation through sentiment analysis,” *Acta Universitatis Sapientiae, Informatica*, vol. 15, no. 2, pp. 330–344, 2023.
- [3] Ramyasri, C. Niharika, K. Maneesh, and B. M. Ismail, “Sentiment analysis of patients’ opinions in healthcare using lexicon-based method,” *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1, pp. 2249–8958, 2019.
- [4] A. Hopper and M. Uriyo, “Using sentiment analysis to review patient satisfaction data located on the internet,” *Journal of Health Organization and Management*, vol. 29, no. 2, pp. 221–233, 2015.
- [5] T. Khan and S. Khalid, “Sentiment analysis for health care,” *International Journal of Privacy and Health Information Management*, vol. 3, no. 2, pp. 78–91, 2015.
- [6] J. Pavlopoulos and I. Androutsopoulos, “Aspect term extraction for sentiment analysis: New datasets, new evaluation measures and an improved unsupervised method,” in *Proc. LASMEACL*, pp. 44–52, 2014.
- [7] T. Wang, Y. Cai, H. F. Leung, R. Y. Lau, Q. Li, and H. Min, “Product aspect extraction supervised with online domain knowledge,” *Knowledge-Based Systems*, vol. 71, pp. 86–100, 2014.
- [8] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi, and L. Donaldson, “Use of sentiment analysis for capturing patient experience from free-text comments posted online,” *Journal of Medical Internet Research*, vol. 15, no. 11, p. e239, 2013.
- [9] M. Qingliang, L. Qiudan, and D. Ruwei, “AMAZING: A sentiment mining and retrieval system,” *Expert Systems with Applications*, vol. 36, pp. 7192–7198, 2009.
- [10] Y. Qiang, Z. Ziqiong, and R. Li, “Sentiment classification of online reviews to travel destinations by supervised machine learning approaches,” *Expert Systems with Applications*, vol. 36, no. 1, pp. 6527–6535, 2009.
- [11] A. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, vol. 1, no. 2, pp. 1–135, 2008.
- [12] S. Sarawagi, “Information extraction,” *Foundations and Trends in Databases*, vol. 1, no. 3, pp. 261–377, 2008.
- [13] R. L. Cilibrasi and P. M. B. Vitányi, “The Google similarity distance,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 3, pp. 370–383, 2007.
- [14] M. Chau and J. Xu, “Mining communities and their relationships in blogs: A study of online hate groups,” *International Journal of Human–Computer Studies*, vol. 65, no. 1, pp. 57–70, 2007.
- [15] A. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet allocation,” *Journal of Machine Learning Research*, vol. 3, no. 1, pp. 993–1022, 2003.

A Study of Sentiment Analysis of Drug Reviews using Transformer-Based Embeddings and Hybrid Classifiers

Nayan Sarkar ¹, Dr. Balveer Singh ²

^{1,2} Department of Computer Science & Engineering, P.K. University, Shivpuri, M.P., India.

ABSTRACT

This study presents a sentiment analysis model for classifying user-generated drug reviews into positive, neutral, and negative categories using transformer-based embeddings and hybrid classifiers. A dataset of 5,170 drug reviews was collected from WebMD using the BeautifulSoup Python library and manually labeled for accuracy. After preprocessing and encoding, embeddings were generated using BERT, SciBERT, BioBERT, and SBERT models to capture contextual meaning. These embeddings were then used as input features for machine learning classifiers such as Decision Tree, Support Vector Machine (SVC), Random Forest, and Recurrent Neural Network (RNN). Experiments were conducted using Python (v3.9.11) and the SMOTE technique to address class imbalance. Performance evaluation using precision, recall, and F1-score revealed that the Random Forest and RNN models achieved the highest accuracy. The results demonstrate that transformer-based embeddings combined with hybrid classifiers offer an effective framework for analyzing public sentiment on drug efficacy and safety.

Keywords: *Sentiment Analysis, Drug Reviews, Recurrent Neural Network (RNN), Machine Learning.*

1. Introduction

There has been an explosion of user-generated information, especially evaluations and experiences from patients about different medications and medical treatments, due to the proliferation of digital health platforms. When it comes to medicine efficacy, safety, and patient happiness, these evaluations are a goldmine of practical information. Nevertheless, there are substantial obstacles to successful analysis of textual feedback due to its unstructured character. One answer is sentiment analysis, a branch of NLP that automates the process of extracting and interpreting feelings and opinions from text. Aiming to enhance healthcare decision-making and patient-centered medication development, this research uses machine learning and deep learning methods to assess user-generated drug evaluations. The goal is to turn unstructured patient input into relevant pharmaceutical insights.

This project aims to construct an all-encompassing sentiment analysis framework that can reliably categorize evaluations of drugs as either favorable, negative, or neutral. The suggested system would decipher the subtleties of language, medical jargon, and emotional undertones in patient accounts by using sophisticated natural language processing models and machine learning techniques. Because it sheds light on medication effectiveness, patient adherence, and safety issues from an evidence-based perspective, understanding these attitudes is vital for pharmaceutical firms, healthcare providers, and lawmakers. Systematically analyzing patient input for improved clinical and commercial decision-making is vital due to the growing confidence patients have in online health communities.

Data gathering and preparation is the first stage of this procedure. Verified medication review sites and internet health discussion groups will be the sources of user-generated content. We will scrape the web using Python's BeautifulSoup package to efficiently retrieve data such as review content, drug name, rating, and date. After extraction, the raw text is preprocessed using natural language processing tools such as NLTK and spaCy to lowercase, tokenize, remove stop words, and lemmatize. In order to prepare the text

for feature extraction and training machine learning models, this step is necessary to clean and standardize it. Improving model performance relies heavily on proper preprocessing, which involves removing noise and highlighting pertinent textual patterns that match to sentiment expressions.

Utilizing word embeddings and contextual language models, the analyzed data is subsequently numerically represented. In many cases, the contextual links between words cannot be captured by traditional approaches such as TF-IDF and Bag-of-Words (BoW). The study gets around these restrictions by using contextualized embeddings generated by pre-trained transformer-based models. As a starting point, one of these models is BERT, which stands for Bidirectional Encoder Representations from Transformers. By taking into account both the left and right word dependencies in a phrase, BERT is able to represent bidirectional context. Because of this, the model can comprehend nuanced language, like denial and sarcasm, which are common in user evaluations. "This drug worked well initially, but caused severe headaches later" is an example of a mixed-sentiment statement that BERT's contextual awareness aids in accurately classifying, as opposed to previous models.

The study also makes use of two specialized versions of the BERT model, SciBERT and BioBERT. SciBERT excels in comprehending domain-specific terms and technical jargon because it is educated on academic and scientific literature. By contrast, BioBERT is very effective at processing healthcare-related terms and medical terminology since it is trained on biomedical corpora like PubMed and clinical notes. By improving the semantic comprehension of drug review medical language, these models increase sentiment classification accuracy. Phrases such as "caused mild nausea" or "reduced systolic pressure" have substantial contextual meaning in biological language, which models like BioBERT can accurately decipher, even if they do not convey express emotional tone.

Also, for semantic analysis at the sentence level, Sentence-BERT (SBERT) is used. Using a Siamese network design, SBERT creates sentence embeddings of predetermined sizes, in contrast to regular BERT that gives embeddings at the token level. Subject similarity analysis, emotion grouping, and paraphrasing identification are all made easier with this capability. Research shows that SBERT is useful for categorizing reviews with comparable sentiments or for spotting small differences in sentiment intensity. A more sophisticated understanding of patient feedback is made possible by this sentence-level depiction, which aids pharmaceutical stakeholders in identifying trends, such as recurrent adverse effects or overall satisfaction patterns across several medications.

After the feature extraction step is over, sentiment classification is done using a variety of machine learning classifiers. To set the stage for baseline performance measurements, classic models include Decision Tree, Support Vector Classification (SVC), and Random Forest. To better comprehend which textual characteristics contribute most to sentiment prediction, Decision Tree classifiers create hierarchical divides based on features that best separate sentiment classes. This makes them clearly interpretable. Conversely, SVC finds the best hyperplane to divide sentiment categories in a space with many dimensions. Complex text data is well-suited to SVC because, with the help of kernel functions, it can effectively manage non-linear connections.

As an ensemble learning technique, Random Forest uses a combination of Decision Trees to make predictions more stable and less prone to overfitting. To make sure the categorization findings are solid, it uses majority voting to combine the output from many trees. Important for dealing with varied and unexpected user-generated evaluations, this ensemble method enables the model to generalize well to unseen data. To further understand the sequential relationships in text data, Recurrent Neural Networks (RNNs) are used. Reasoning neural networks (RNNs) are useful for simulating the temporal sentiment

changes that occur in drug evaluations, which often include numerous words with changing sentiments over time (e.g., early pleasure followed by unfavorable experiences). The network is able to learn context from word to word thanks to its recurrent connections, which enable it to remember what words have come before.

The research uses stringent assessment standards to guarantee the findings are reliable. To test the success of the classification, the models are evaluated using metrics like as accuracy, precision, recall, F1-score, and ROC-AUC. Also, to make sure the model is consistent and to reduce bias, k-fold cross-validation is run. One way to see how well the models do at classifying attitudes as positive, negative, or neutral is via a confusion matrix. To find the best mix of algorithms and embedding models for sentiment classification, it is necessary to compare them.

Various open-source libraries and frameworks are essential to the execution of this study. Because of its great support for NLP and machine learning applications, Python is the main language used for programming. For traditional methods, you may use scikit-learn or another library, and for deploying deep learning models, you can use TensorFlow or PyTorch. Hugging Face's Transformers library streamlines the process of fine-tuning models for sentiment analysis by providing access to pre-trained models such as BERT, SciBERT, BioBERT, and SBERT. Matplotlib and Seaborn are used for data visualization and performance analysis. These tools make it easier to understand results and display conclusions visually.

2. Review of Literature

Alkhnabshi, Omer et al., (2024) In recent years, internet medical forums have grown into a vital, low-cost resource for medical service management input, where patients may share their stories and get answers from others. This data provides insight into the efficacy of pharmaceutical treatments, pain management tactics, and alternative therapies, in addition to helping to gauge patient happiness and enhance healthcare quality. This study thoroughly documents key aspects of patients' experiences, showcasing the positive and negative emotions expressed in their narratives. Our dataset includes around 15,000 items extracted from various sections of the well-known patient.info medical forum. Using a unique combination of content analysis, deep learning techniques, aspect-based sentiment analysis, and LLM, we conduct an examination of this data. The goal of our method is to unearth the many facets conveyed by patients' comments. The experiment verified that deep learning models could properly predict the sentiment values linked with seven distinct aspect categories found in the feedback. The LLM with few-shot learning stood out from the rest of the models. Our study provides valuable insights into patients' experiences in online forums and highlights the usefulness of advanced analytical methods for extracting meaning from patients' unstructured input. This knowledge might be very beneficial for healthcare professionals and medical service management.

Tanasescu, Laura et al., (2024) It is critical for most companies to increase staff performance and productivity because of the various changes that have occurred in people and processes in many organizations in the last few years. The fact that this can keep the business growing and competitive makes it even more important. Creating a machine learning system to predict the performance ratings of people in a particular company entails data pre-processing, variable selection, algorithm building, and hyperparameter shaping. Various methods are compared and contrasted in this research. In order to increase overall productivity and decrease reliance on subjective human judgment in performance assessments, we set out to determine the most effective strategies for predicting the provided factors.

Raju, Nidadavolu et al., (2024) Healthcare, banking, and e-commerce decision-makers are seeking relevant insights from big data, and predictive analytics is a vital tool for them due to the ever-increasing complexity and volume of data. When working with large amounts of data, it may be challenging to choose the optimal machine learning algorithm due to differences in processing requirements, scalability, and algorithmic performance. This paper provides a comprehensive review of popular machine learning algorithms for predictive analytics, focusing on their performance and practicality in big data environments. Research sorts algorithms into three categories based on their learning type: supervised, unsupervised, and reinforcement learning. Algorithms are then evaluated based on many factors, such as the accuracy of their predictions, the efficiency of their computations, their scalability, and their suitability for real-time analytics. In order to determine how well they manage massive datasets, a number of techniques are tested and evaluated, including decision trees, linear regression, neural networks, support vector machines, and clustering methods. The study offers a set of metrics, including training time, F1-score, and accuracy, to further assess the algorithms' computational feasibility and predictive potential. Through the application of these algorithms to a sample of a huge data collection, this case study sheds light on their real-world performance in several predictive analytics scenarios. Comparison tables and performance graphs are good visual representations of the data that assist to comprehend the pros and cons of various algorithm possibilities. Algorithms like random forests and neural networks perform better at generating predictions, but they may be too resource-intensive for real-time processing in big data applications, according to the study. In this work's later half, we provide suggestions for selecting machine learning algorithms based on specific predictive analytics objectives, data attributes, and processing requirements. The article continues by discussing the challenges of implementing these algorithms in big data environments, but it also offers solutions, such as incorporating deep learning and employing distributed computing, that might enhance predictive analytics going ahead.

Udeh, Chioma et al., (2024) Big Data Analytics is revolutionizing the way businesses get insights from diverse and expansive datasets in the dynamic realm of business intelligence. After a lengthy analysis of Big Data Analytics and its impact on modern BI, this paper provides a concise summary of the key issues. Big Data Analytics represents a sea shift in how organizations make decisions by allowing them to fully use their data assets. The study delves into the many uses of Big Data Analytics, showcasing its significance in customer-centric endeavors, operational effectiveness, risk mitigation, and long-term strategy formulation. Companies may greatly enhance their strategic planning with the help of predictive analytics when they utilize it to anticipate market shifts. Incorporating analytics-derived insights into decision-making drives future-oriented strategic efforts that are better informed. When it comes to identifying and preventing fraud, risk management becomes much more proactive with the integration of Big Data Analytics. The ability to manage enormous volumes of data in real-time raises awareness while decreasing financial risks associated with fraudulent activities. Through the use of scenario modeling, organizations may improve their risk mitigation and anticipation capabilities. The focus shifts to operational efficiency as analytics expose inefficiencies in supply chains, retail operations, and industrial processes. Through the use of data analytics, which enable real-time decision-making, retailers can maintain a responsiveness to changing customer preferences and market circumstances. Increased focus on customers is a direct result of personalized marketing and predictive analytics used by customer care departments. In an effort to boost customer satisfaction and retention, this analysis looks at how Big Data Analytics enables companies to provide personalized experiences. The paper provides a concise overview of Big Data Analytics' game-changing trajectory in modern BI, focusing on its role in simplifying strategic complexity, decreasing risk, enhancing operations, and making customer-centric choices. In today's data-driven world, businesses may benefit from a comprehensive review that tries to show them how to use Big Data Analytics to their advantage.

Hanif, Atikah et al., (2023) Technological advancements have caused an explosion of data in several industries, including aviation. With the use of big data analytics, it is possible to extract valuable insights from these enormous datasets. Relevant papers were selected from twenty academic articles that were indexed in internet databases such as ScienceDirect, Web of Science, Google Scholar, and IEEE Xplore. The words "data mining," "machine learning," "big data analytics," "predictive analytics," and the "airlines industry" were all part of the search. Based on the findings, Big Data Analytics (BDA) offers several applications in the aviation industry, including optimizing airline operations, increasing customer pleasure and service quality, strengthening safety and risk management, and aircraft maintenance. The investigation reveals that there are many significant challenges to the adoption of BDA in the aviation industry, including concerns over data integration, privacy and security, and compliance with regulations. When it comes down to it, this review article is a gold mine of information on the pros and cons of BDA in the aviation industry. In doing so, it increases our understanding of the subject and prepares the way for future research and practical applications.

3. Proposed Approach

Libraries and Models Used

The libraries that were used and the many pre-trained models are listed below: -

- BeautifulSoup
- Bidirectional Encoder Representations from Transformers
- SciBERT
- BioBERT
- SBERT
- Decision Tree
- Support Vector Classification
- Random Forest
- Recurrent Neural Network

Dataset

The dataset was constructed by manually classifying 5,170 medication reviews into three separate categories according to their sentiment using the BeautifulSoup Python package. The reviews were sourced from the publically available website WebMD. The dataset included two columns, "Reviews" and "Classification," since the categories were "Positive," "Neutral," and "Negative."

Preprocessing

The first and most important stage in data mining and machine learning is data preprocessing, which entails cleaning and organizing raw data so it can be better analyzed and modelled. Data accuracy, consistency, and completeness are ensured by a sequence of procedures that deal with discrepancies, missing numbers, outliers, and mistakes. In order to improve the data's quality and make it more usable for different algorithms, data preparation also include data reduction, data transformation, and feature engineering.

Generation of Embeddings

We need to convert the pre-processed data, which includes reviews and labels, into numerical numbers. This led to the creation of embeddings of the reviews and the ability to encode labels using LabelEncoder. A low-dimensional representation of a high-dimensional data set, such text, pictures, or sounds, is called an embedding. Automated systems are able to comprehend and handle complicated data because they record the

significance and connections between the data elements. Embeddings are used in several applications, such as recommender systems, image recognition, anomaly detection, and natural language processing, and they are trained using massive volumes of data. A number of natural language processing models, such as S-BERT, BioBERT, SciBERT, and BERT, are fed the reviews. By taking the review's context into account, the embeddings acquired from these models capture the text's meaning as vectors representing the reviews.

Proposed Models of Machine Learning

The categorization procedure starts when the embeddings are generated and their labels are identified. This is done by using different machine learning models to process the embeddings along with the labels that go with them. Decision tree, logistic regression, support vector, and random forest classifiers are some of the most popular machine learning methods used for classification problems. There are advantages and disadvantages to each of these algorithms; picking the right one depends on the details of the data and the result you're after.

Model Comparison

We compare and assess the models' performance. Since RNN produces better results in terms of accuracy, it is selected for use in model construction based on evaluation criteria.

Experimental Procedure

The machine running Windows 11 and powered by an Intel Core i5 CPU was used to complete this job. Visual Studio Code in conjunction with the Python 3.9.11 environment was used to run the code. Since gathering data is the first order of business, web scraping should be executed before a plethora of other responsibilities. To get over the issue of dataset imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is used. A dataset with ratings and labels is now available. We used LabelEncoder to encode the labels. The reviews were embedded using several pre-trained models such as Bert, SciBERT, BioBERT, and SBERT. The Python Transformers module was used for this purpose. Training and testing sets are subsequently created from the dataset. A model's generalizability and prediction power on novel and unseen data may be better evaluated with correct data splitting. Two variables were used: one for storing independent characteristics and another for saving goal values. The dataset was analyzed using a number of classification models, such as Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest. We need to choose one of them for our model based on this. Among the classifiers, Random Forest has the highest accuracy. To improve accuracy and decrease variation, an ensemble learning approach called a random forest classifier uses many decision trees to produce predictions.

4. Results and Discussion

In this part, we provide the results of the experiment that used the suggested method for sentiment analysis on UGC on drugs. In this part, we provide the comparison findings and then go into depth about what we found.

Table 1: Performance Metrics for Various Classifiers using BERT with Recurrent Neural Network

Sentiment	Precision	Recall	F1-Score	Support
Negative	0.63	0.70	0.66	510
Neutral	0.40	0.78	0.16	205
Positive	0.47	0.60	0.53	325
Accuracy			0.55	1040
Macro Avg	0.50	0.46	0.45	1040
Weighted Avg	0.54	0.55	0.52	1040

Table 1 presents the performance metrics of a Recurrent Neural Network (RNN) model using BERT embeddings for sentiment analysis of drug reviews. The model demonstrates the highest recall for neutral sentiments (0.78), indicating that it is effective at identifying most neutral reviews, although the corresponding F1-score is low (0.16), suggesting imbalanced prediction performance for this class. Negative sentiment shows a balanced performance, with a precision of 0.63, recall of 0.70, and an F1-score of 0.66, reflecting the model's relatively strong ability to correctly classify negative reviews. Positive sentiment is moderately predicted, with precision at 0.47, recall at 0.60, and F1-score at 0.53, showing that some positive reviews are misclassified as other sentiments.

Table 2: Performance Metrics for Various Classifiers using SBERT with Recurrent Neural Network

Sentiment	Precision	Recall	F1-Score	Support
Negative	0.64	0.74	0.68	510
Neutral	0.29	0.23	0.25	205
Positive	0.52	0.46	0.48	325
Accuracy			0.55	1040
Macro Avg	0.48	0.47	0.47	1040
Weighted Avg	0.53	0.55	0.53	1040

Table 2 displays the performance metrics of a Recurrent Neural Network (RNN) model using SBERT embeddings for sentiment analysis of drug reviews. The model shows its strongest performance in detecting negative sentiments, with a precision of 0.64, recall of 0.74, and F1-score of 0.68, indicating that negative reviews are identified accurately and consistently. Positive sentiment classification is moderate, with precision at 0.52, recall at 0.46, and an F1-score of 0.48, suggesting that some positive reviews are misclassified as neutral or negative. The neutral class demonstrates weaker performance, with a precision of 0.29, recall of 0.23, and F1-score of 0.25, indicating difficulty in correctly identifying neutral reviews.

Table 3: Performance Metrics for Various Classifiers using SciBERT with Recurrent Neural Network

Sentiment	Precision	Recall	F1-Score	Support
Negative	0.58	0.77	0.66	510
Neutral	0.35	0.10	0.16	205
Positive	0.49	0.48	0.48	325
Accuracy			0.56	1040
Macro Avg	0.47	0.45	0.43	1040
Weighted Avg	0.51	0.55	0.50	1040

Table 3 presents the performance of a Recurrent Neural Network (RNN) model using SciBERT embeddings for sentiment analysis of drug reviews. The model exhibits strong performance in detecting negative sentiments, achieving a precision of 0.58, recall of 0.77, and F1-score of 0.66, indicating effective identification of negative reviews. Positive sentiment classification is moderate, with precision at 0.49, recall at 0.48, and an F1-score of 0.48, showing balanced but not optimal performance. Neutral sentiment detection is relatively weak, with a precision of 0.35, recall of 0.10, and F1-score of 0.16, highlighting the model's difficulty in correctly classifying neutral reviews.

Table 4: Performance Metrics for Various Classifiers using BioBERT with Recurrent Neural Network

Sentiment	Precision	Recall	F1-Score	Support
Negative	0.62	0.73	0.67	510
Neutral	0.32	0.12	0.18	205
Positive	0.50	0.55	0.52	325
Accuracy			0.56	1040
Macro Avg	0.48	0.47	0.46	1040
Weighted Avg	0.52	0.56	0.53	1040

Table 4 shows the performance of a Recurrent Neural Network (RNN) model using BioBERT embeddings for sentiment analysis of drug reviews. The model demonstrates strong performance in detecting negative sentiments, with a precision of 0.62, recall of 0.73, and F1-score of 0.67, indicating reliable classification of negative reviews. Positive sentiment is moderately well predicted, with precision at 0.50, recall at 0.55, and F1-score at 0.52, suggesting that the model can capture positive sentiments reasonably accurately. Neutral sentiment detection remains challenging, with a precision of 0.32, recall of 0.12, and F1-score of 0.18, highlighting difficulty in correctly identifying neutral reviews.

5. Conclusion

Artificial intelligence has the ability to revolutionize the pharmaceutical and healthcare industries, as shown by research that used machine learning methods to analyze user-generated medication evaluations sentiment. The study shows that patient views may be systematically analyzed to find out about treatment effectiveness, side effects, and overall satisfaction using sophisticated natural language processing (NLP) models including BERT, SciBERT, BioBERT, and SBERT. By combining several machine learning methods such as Support Vector Classification, Random Forest, and Recurrent Neural Networks, sentiment classification becomes more accurate and dependable. This guarantees that positive, negative, and neutral feedback are properly distinguished.

By taking this approach, lawmakers, healthcare providers, and pharmaceutical firms may all make evidence-based choices that put patients' needs first. Product enhancement, patient education, and post-market monitoring may all benefit from the results of these types of research. Furthermore, stakeholders may enhance the healthcare ecosystem's responsiveness, openness, and trust by comprehending public mood patterns. Finally, the study proves that sentiment analysis is a must-have for future patient-centered healthcare by helping to close the information gap between patients and the pharmaceutical sector.

References

1. Hanif and A. Jaafar, "A Review: Application of Big Data Analytics in Airlines Industry," *Open International Journal of Informatics*, vol. 11, no. 2, pp. 196-209, 2023, doi: 10.11113/oiji2023.11n2.254.
2. C. Udeh, O. Orieno, O. Daraojimba, N. Ndubuisi, and O. Oriekhoe, "Big Data Analytics: A Review of Its Transformative Role in Modern Business Intelligence," *Computer Science & IT Research Journal*, vol. 5, no. 1, pp. 219-236, 2024, doi: 10.51594/csitrj.v5i1.718.
3. N. Raju and P. Devi, "A Comparative Analysis of Machine Learning Algorithms for Big Data Applications in Predictive Analytics," *International Journal of Scientific Research and Management (IJSRM)*, vol. 12, no. 10, pp. 1608-1630, 2024, doi: 10.18535/ijssrm/v12i10.ec09.
4. L. Tanasescu, A. Vines, A.-R. Bologa, and O. Virgolici, "Data Analytics for Optimizing and Predicting Employee Performance," *Applied Sciences*, vol. 14, no. 1, pp. 3254, 2024, doi: 10.3390/app14083254.

5. O. Alkhnbashi, R. Nassr, and M. Hammoudeh, "Aspect-Based Sentiment Analysis of Patient Feedback Using Large Language Models," *Big Data and Cognitive Computing*, vol. 8, no. 12, pp. 2-29, 2024, doi: 10.3390/bdcc8120167.
6. T. V. Iyelolu, E. E. Agu, C. Idemudia, and T. I. Ijomah, "Leveraging artificial intelligence for personalized marketing campaigns to improve conversion rates," *International Journal of Engineering Research and Development*, vol. 20, no. 8, pp. 253–270, 2024.
7. M. R. Ayyagari, "A framework for analytical CRM assessments challenges and recommendations," *International Journal of Information, Business and Management*, vol. 13, no. 2, pp. 108–121, 2021.
8. K. W. Walker, "Exploring adaptive boosting (AdaBoost) as a platform for the predictive modeling of tangible collection usage," *The Journal of Academic Librarianship*, vol. 47, no. 6, p. 102450, 2021.
9. Sumit, "Drug review sentiment analysis using boosting algorithms," *International Journal of Trend in Scientific Research and Development (IJTSRD)*, vol. 5, no. 4, pp. 937–941, 2021.
10. V. Djuriscic, L. Kascelan, and S. Rogic, "Bank CRM optimization using predictive classification based on the support vector machine method," *Applied Artificial Intelligence*, vol. 34, no. 1, pp. 1–22, 2020.
11. B. Zorić, "Benefits of educational data mining," *Journal of International Business Research and Marketing*, vol. 6, no. 1, pp. 12–16, 2020.
12. Y. Rohani, Z. Torabi, and S. Kianian, "A novel hybrid genetic algorithm to predict students' academic performance," *Journal of Electrical and Computer Engineering Innovations*, vol. 8, no. 2, pp. 219–232, 2020.
13. F. Ofori, E. Maina, and R. Gitonga, "Using machine learning algorithms to predict students' performance and improve learning outcomes: A literature-based review," *Journal of Information and Technology*, vol. 4, no. 1, pp. 33–55, 2020.
14. D. Buenaño-Fernández, D. Gil, and S. Luján-Mora, "Application of machine learning in predicting performance for computer engineering students: A case study," *Sustainability*, vol. 11, no. 10, pp. 2833, 2019.
15. R. P. Bunker and F. Thabtah, "A machine learning framework for sport result prediction," *Applied Computing and Informatics*, vol. 15, no. 1, pp. 27–33, 2019.
16. N. Chen, C. Xiong, W. Du, C. Wang, X. Lin, and Z. Chen, "An improved genetic algorithm coupling a back-propagation neural network model (IGA-BPNN) for water-level predictions," *Water*, vol. 11, no. 9, p. 1795, 2019.
17. Sahoo et al., "Deep learning-based health recommender system using collaborative filtering," *Computation*, vol. 7, no. 2, pp. 1283–1299, 2019.
18. V. Deshmukh et al., "Comparative study of different approaches of opinion mining," *International Journal of Research*, vol. 8, no. 1, pp. 644–651, 2019.
19. M. Abdullah and A. Fahad, "Enhancing performance of educational data using big data and Hadoop," *International Journal of Applied Engineering Research*, vol. 14, no. 19, pp. 3814–3819, 2019.
20. Al-Homery, H. Asharai, and A. Ahmad, "The core components and types of CRM," *Pakistan Journal of Humanities and Social Sciences*, vol. 7, no. 1, pp. 121–145, 2019.
21. E. T. Lau, L. Sun, and Q. Yang, "Modeling, prediction, and classification of student academic performance using artificial neural networks," *SN Applied Sciences*, vol. 1, no. 9, pp. 1–10, 2019.