

A Review of Data Analytics Techniques Used to Analyze Patient Satisfaction in Healthcare

Praveen Soneja¹, Anamika Tiwari², Anand Kumar, Aniket Kumar, Animesh Raj

¹ Noida Institute of Engineering & Technology (MCA Institute), Greater Noida

² Noida Institute of Engineering & Technology (MCA Institute), Greater Noida

Abstract: Advertising plays a crucial role in influencing customer purchasing behaviour and improving business revenue. This study analyses the relationship between advertising expenditure on three media channels—TV, Radio, and Newspaper—and product sales using a supervised machine learning approach. A linear regression model is developed to predict sales performance based on advertising budgets. The dataset consists of 200 observations and four variables. Exploratory data analysis, model training, performance evaluation, and visualization techniques are applied. The results indicate that TV and Radio advertisements have a stronger impact on sales compared to Newspaper advertisements. The proposed model demonstrates good prediction accuracy and provides useful insights for marketing decision-making.. It looks at popular data sources, analytical approaches from machine learning and natural language processing to conventional statistical methods, and the major variables affecting patient satisfaction. Along with discussing current issues and potential future research topics, the report also evaluates the effectiveness, benefits, and drawbacks of various analytical methodologies. The results are intended to assist politicians, researchers, and medical professionals in using data-driven approaches to improve patient-centered healthcare delivery.

Keywords — Patient Satisfaction, Patient Experience, Health-care Analytics, Data Analytics Techniques, Machine Learning.

I. INTRODUCTION

It is well acknowledged that patient happiness is a key measure of the effectiveness and quality of healthcare. It represents how patients see the efficacy of their care, the quality of their communications, the availability of resources, and their general healthcare experience. In the past, clinical outcomes like death rates, infection rates, and treatment efficacy were the main emphasis of healthcare evaluation. However, these metrics by themselves are unable to fully represent patients' subjective experiences, which are vital in judging the effectiveness of treatment. Patient satisfaction has become a crucial metric for evaluating organizational efficacy and service quality as healthcare organizations progressively embrace patient-centered care models.

1.1 Background and Context

Over time, the idea of patient satisfaction has undergone substantial change. Patient satisfaction was formerly seen as a secondary objective and was frequently assessed using straightforward feedback questionnaires or complaint logs. It has evolved over the last few decades into a regulated and organized quality indicator that is utilized by legislators, accreditation authorities, and healthcare organizations. The significance of measuring patient happiness has been further highlighted by the growing emphasis on accountability, transparency, and value-based care. The focus of healthcare research has recently switched from patient satisfaction to patient experience.

Patient experience encompasses the full care journey, including encounters with medical personnel, emotional support, care coordination, and ease of access to services, whereas patient satisfaction assesses whether patient expectations are fulfilled. This distinction is crucial since patient experience offers greater insights into certain areas that need to be improved. Research has shown that better clinical outcomes, better treatment plan adherence, lower hospital readmission rates, greater patient loyalty, improved organizational reputation, and financial sustainability are all linked to higher patient satisfaction and positive patient experiences.

1.2 The Role of Data Analytics in Healthcare

Electronic health records (EHRs), hospital information systems, wearable technology, mobile health apps, and online healthcare platforms are all widely used as a result of the quick digital transformation of the healthcare industry. Every day, these technologies produce enormous amounts of data. Both structured and unstructured methods—such as free-text comments, patient narratives, social media posts, and online reviews—are used to gather patient satisfaction data. Standardized surveys, questionnaires, and rating scales are examples of organized formats.

The size, complexity, and variety of such data frequently make traditional descriptive and summary-based analysis techniques inadequate. Healthcare firms may now use predictive and prescriptive analytics instead of only descriptive analysis because to developments in data analytics. To find trends, forecast satisfaction levels, and discover important factors influencing patient views, statistical analysis, machine learning algorithms, and natural language processing approaches are being employed more and more. Even if there is a lot of data accessible, there is still a big gap between data gathering and the creation of useful insights that may successfully raise patient happiness and healthcare quality.

1.3 Research Questions

The following research questions are the focus of this review paper: • RQ1: Which data sources and techniques are frequently employed to get information on patient satisfaction?

- RQ2: Which data analytics methods have been used to examine patient satisfaction in the medical field?
- RQ3: What important variables and indicators of patient satisfaction have been found using data analytics?
- RQ4: How do the performance, benefits, and drawbacks of various analytical techniques compare?
- RQ5: What are the key obstacles, unmet research needs, and potential avenues for further study in this field?

1.4 Scope and Definition

In this study, patient experience refers to the wider range of patient contacts and impressions over the course of care, whereas patient satisfaction refers to patients' subjective assessment of healthcare services. Statistical analysis, machine learning models, natural language processing, and predictive analytics applied to patient satisfaction data are all included in the category of data analytics approaches. The study takes into account research from a variety of healthcare environments, such as clinics, hospitals, and specialty care centers. In order to capture current trends and developing difficulties in patient satisfaction analytics, the temporal scope concentrates on recent and significant research.

METHODOLOGY

In order to find, evaluate, and compile pertinent literature on data analytics methods used to assess patient satisfaction in healthcare, this study employs a methodical methodology. For the purpose of gathering literature, academic databases including IEEE Xplore, Scopus, PubMed, SpringerLink, and Google Scholar were taken into consideration. Relevant studies were found using keywords such as "patient satisfaction," "patient experience," "healthcare analytics," "machine learning in healthcare," and "sentiment analysis in healthcare."

Relevance, publication quality, and contribution to patient satisfaction analysis were the criteria used to screen the chosen research. To give a thorough review of analytical techniques, both quantitative and qualitative investigations were covered. Data sources, analytical methods, and application areas were used to classify the evaluated literature. This approach guarantees a methodical and objective summary of current research.

Patient Satisfaction Data Sources

Analysis of patient satisfaction is highly dependent on the caliber and variety of data sources. Patient satisfaction statistics are now gathered from a variety of sources due to the growing use of digital health technology, allowing for a more thorough comprehension of patient experiences and perspectives. These sources of information may be roughly divided into two categories: organized and unstructured.

3.1 Structured Data Sources

The most popular and well recognized methods of measuring patient satisfaction are structured data sources. These consist of rating scales, questionnaires, and standardized surveys intended to quantify patient input. Healthcare organizations frequently employ well-known tools like the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey and customized hospital feedback forms. Likert-scale items about staff responsiveness, waiting times, cleanliness, communication with healthcare providers, and general satisfaction are often included in these surveys. Because of their predetermined format, structured data are comparatively simple to examine using statistical and machine learning approaches. They make it possible to compare between hospitals, departments, and time periods. However, because replies are limited, such data could not fully convey the complexity and variety of patients' feelings and experiences. However, because replies are limited by predetermined possibilities, such data may fall short of capturing the richness and complexity of patient feelings and experiences.

3.2 Unstructured Data Sources

Due to their capacity to offer deeper and more comprehensive insights on patient experiences, unstructured data sources have drawn increased attention in recent years. These sources include social media posts, blogs, discussion forums, online reviews on healthcare platforms, patient complaints, and free-text survey comments. Patients may share their thoughts, feelings, and experiences in their own words using unstructured data, which provides important context that conventional surveys could miss. The amount, diversity, and absence of a consistent format of unstructured data make its interpretation extremely difficult. To extract useful information from textual data, sophisticated methods like sentiment analysis and natural language processing are needed. Despite these difficulties, unstructured data has shown itself to be quite useful for spotting new problems, figuring out patient attitude, and finding previously unknown elements affecting satisfaction.

3.3 Hybrid and Integrated Data Sources

In order to provide a comprehensive picture of patient satisfaction, recent research emphasizes the integration of structured and unstructured data sources. More precise and context-aware analysis is made possible by combining survey results with textual comments and clinical information from electronic health records. Researchers may make better decisions by correlating patient perceptions with clinical outcomes, service quality measures, and operational performance using hybrid datasets.

Data Analytics Techniques for Patient Satisfaction Analysis

Data on patient satisfaction has been analyzed using a variety of data analytics methodologies. The complexity, interpretability, and prediction value of these methods vary, and they are frequently chosen according to the type of data and the goals of the study.

4.1 Descriptive and Statistical Analysis

The fundamental method for analyzing patient satisfaction is descriptive statistics. Survey data are frequently summarized and correlations between satisfaction levels and service features are found using methods including mean scores, frequency distributions, regression models, and correlation analysis. Because statistical approaches are transparent and easy to understand, they may be used for reporting and benchmarking. However, the capacity of conventional statistical methods to handle massive amounts of unstructured data and simulate intricate, nonlinear connections is constrained. More sophisticated analysis techniques are becoming more and more necessary as healthcare databases continue to expand in size and complexity.

4.2 Machine Learning Techniques

Because machine learning approaches can handle high-dimensional data and find intricate patterns, they have become increasingly popular in patient satisfaction assessments. Decision trees, support vector machines, logistic regression, random forests, and neural networks are examples of supervised learning algorithms that are frequently used to categorize patient feedback and forecast patient satisfaction levels. To find common themes in feedback data and to group patients with similar satisfaction profiles, unsupervised learning techniques like topic modeling and clustering are used. When compared to conventional statistical methods, machine learning models frequently show greater predicted accuracy; nevertheless, their interpretability may be compromised, which may restrict their use in clinical decision-making situations.

4.3 Natural Language Processing and Sentiment Analysis

Analyzing unstructured patient feedback requires the use of natural language processing (NLP) tools. Sentiment analysis, which categorizes comments as positive, negative, or neutral, is frequently used to ascertain the polarity of patient attitudes. Key topics, service qualities, and emotional expressions may be extracted from textual data using more sophisticated NLP approaches like topic modeling, named entity identification, and deep learning-based language models. NLP-based methods offer more profound insights into patient feelings and perceptions than are possible with only numerical evaluations. However, issues including ambiguous language, domain-specific terminology, and data privacy concerns continue to be major barriers to wider use.

4.4 Predictive and Prescriptive Analytics

Predictive analytics focuses on forecasting patient satisfaction outcomes based on historical data and identified predictors. These models enable healthcare organizations to proactively address potential dissatisfaction and improve service quality. Prescriptive analytics extends this capability by recommending specific actions or interventions to enhance patient satisfaction. Although still emerging, prescriptive analytics holds significant potential for supporting data-driven decision-making in patient-centered healthcare.

Comparative Analysis of Data Analytics Techniques

The performance, interpretability, data needs, and application of various data analytics methodologies used for patient satisfaction analysis differ. The simplicity, openness, and ease of understanding of traditional statistical approaches are highly valued. They work especially well for comparing satisfaction ratings between departments or organizations and for analyzing structured survey data. Nevertheless, these approaches frequently have trouble simulating nonlinear connections and interactions between several variables. Conversely, machine learning methods show better scalability and predictive performance. High-dimensional datasets and intricate patterns in patient satisfaction data may be handled by algorithms like random forests, support vector machines, and neural networks. These methods work particularly well for analyzing big datasets with both clinical and non-clinical factors. Despite their advantages, machine learning models sometimes lack interpretability, which might hinder healthcare professionals' confidence and adoption. Natural language processing methods are particularly good at drawing conclusions from unstructured textual data, such internet reviews and patient comments. Patients' feelings, worries, and expectations can be better understood using sentiment analysis and topic modeling. However, in order to produce dependable results, NLP techniques need significant computing resources and domain-specific customisation.

Challenges and Limitations

The efficient use of data analytics methods for patient satisfaction analysis is hampered by a number of issues, despite notable advancements. Because patient feedback data frequently contains biases, inconsistencies, and missing numbers, data quality is still a significant problem. Online reviews may reflect extreme viewpoints rather than typical patient experiences, and survey-based statistics may suffer from poor response rates. Significant obstacles may come from ethical and privacy issues. Strict adherence to data privacy laws is necessary since patient satisfaction data is frequently connected to private medical information. It is difficult to guarantee data confidentiality and anonymity while preserving analytical value.

The inability of sophisticated analytics models to be interpreted is another significant drawback. To foster trust and accountability, healthcare decision-makers frequently need models that are clear and understandable. Additionally, the generalizability of analytical models across various contexts is limited by variations in healthcare environments, patient demographics, and cultural variables.

Table I presents a comparative overview of commonly used data analytics techniques for patient satisfaction analysis.

Table-1

COMPARISON OF DATA ANALYTICS TECHNIQUES FOR PATIENT SATISFACTION ANALYSIS

Technique	Data Type	Advantages	Limitations
Statistical Analysis	Structured	Simple, interpretable, easy benchmarking	Limited scalability, weak nonlinear modeling
Machine Learning	Structured and Hybrid	High predictive accuracy, handles large datasets	Low interpretability, requires large training data
Natural Language Processing	Unstructured	Captures patient sentiment and themes	Computationally intensive, language ambiguity

Predictive Analytics	Hybrid	Enables proactive decision-making	Model complexity, data dependency
Prescriptive Analytics	Hybrid	Action oriented recommendations	Limited real-world implementation

Future Research Directions

The development of machine learning models that strike a compromise between accuracy and transparency should be the main goal of future patient satisfaction analytics research. A more thorough picture of patient satisfaction may be obtained by integrating multimodal data, such as clinical outcomes, operational metrics, and patient-generated data.

Another possible path is the use of artificial intelligence-driven decision support systems and real-time analytics. Proactive interventions to enhance patient experiences during the course of treatment, as opposed to after discharge, can be made possible by these technologies. Incorporating ethical frameworks, fairness, and bias mitigation strategies into analytical models should also receive more attention.

These studies can offer important information about the long-term effects of patient-centered programs and policy modifications.

Conclusion

A key factor in evaluating the quality of healthcare is patient happiness, which is also essential for enhancing patient-centered treatment. There are now many chances to use advanced data analytics approaches to assess patient satisfaction due to the growing availability of healthcare data. This review study assessed the advantages and disadvantages of different methodologies while looking at data sources, analytical methods, and important factors affecting patient satisfaction.

There are still issues with data quality, interpretability, and privacy even if machine learning and natural language processing algorithms have strong capabilities for drawing conclusions from complicated information. Realizing the full potential of data-driven patient satisfaction analysis will require addressing these issues and developing analytical techniques. The review's conclusions lay the groundwork for further investigation and encourage the use of analytics-driven tactics to improve patient satisfaction and healthcare service quality.

REFERENCES

- [1] A. DONABEDIAN, "THE QUALITY OF CARE: HOW CAN IT BE ASSESSED?" JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION, VOL. 260, NO. 12, PP. 1743–1748, 1988.
- [2] C. JENKINSON, A. COULTER, S. BRUSTER, N. RICHARDS, AND T. CHANDOLA, "PATIENTS' EXPERIENCES AND SATISFACTION WITH HEALTH CARE: RESULTS OF A QUESTIONNAIRE STUDY OF SPECIFIC ASPECTS OF CARE," QUALITY AND SAFETY IN HEALTH CARE, VOL. 11, NO. 4, PP. 335–339, 2002.
- [3] D. R. CANO, J. C. A. GARCÍA, AND J. M. S. GÓMEZ, "PATIENT SATISFACTION IN HEALTHCARE SERVICES: A SYSTEMATIC REVIEW," HEALTH POLICY, VOL. 122, NO. 6, PP. 1–10, 2018.
- [4] CENTERS FOR MEDICARE & MEDICAID SERVICES, "HOSPITAL CONSUMER ASSESSMENT OF HEALTHCARE PROVIDERS AND SYSTEMS (HCAHPS)," 2023.
- [5] S. R. SAFRAN ET AL., "LINKING PATIENT SATISFACTION WITH QUALITY OF CARE," ANNALS OF INTERNAL MEDICINE, VOL. 138, NO. 1, PP. 42–48, 2003.
- [6] J. HAN, M. KAMBER, AND J. PEI, DATA MINING: CONCEPTS AND TECHNIQUES, 3RD ED., MORGAN KAUFMANN, 2012.
- [7] E. H. SHORTLIFFE AND J. J. CIMINO, BIOMEDICAL INFORMATICS: COMPUTER APPLICATIONS IN HEALTH CARE AND BIOMEDICINE, 4TH ED., SPRINGER, 2014.
- [8] I. GOODFELLOW, Y. BENGIO, AND A. COURVILLE, DEEP LEARNING, MIT PRESS, 2016.
- [9] B. LIU, SENTIMENT ANALYSIS AND OPINION MINING, MORGAN & CLAYPOOL PUBLISHERS, 2012.

- [10] S. GEETHA, A. SINGHA, AND S. SINHA, "RELATIONSHIP BETWEEN CUSTOMER SENTIMENT AND ONLINE CUSTOMER RATINGS FOR HOTELS," *TOURISM MANAGEMENT*, VOL. 61, PP. 43–54, 2017.
- [11] M. GHASSEMI, L. A. CELI, AND D. J. STONE, "STATE OF THE ART REVIEW: DATA MINING AND PREDICTIVE ANALYTICS IN HEALTHCARE," *CRITICAL CARE*, VOL. 19, NO. 1, PP. 1–11, 2015.
- [12] H. K. KIM, J. Y. LEE, AND S. J. LEE, "PREDICTING PATIENT SATISFACTION USING MACHINE LEARNING TECHNIQUES," *INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS*, VOL. 128, PP. 50–58, 2019.
- [13] R. FELDMAN, "TECHNIQUES AND APPLICATIONS FOR SENTIMENT ANALYSIS," *COMMUNICATIONS OF THE ACM*, VOL. 56, NO. 4, PP. 82–89, 2013.
- [14] J. BATES, "HEALTH CARE ANALYTICS: DEFINING DATA-DRIVEN DECISION SUPPORT," *HEALTH INFORMATICS JOURNAL*, VOL. 21, NO. 3, PP. 1–9, 2014.
- [15] WORLD HEALTH ORGANIZATION, "QUALITY OF CARE: A PROCESS FOR MAKING STRATEGIC CHOICES IN HEALTH SYSTEMS," WHO PRESS, 2006.