**Chapter 1: Introduction**

**1.Background**

In today’s rapidly evolving business landscape, customer satisfaction is not only a measure of success but also a predictor of long-term viability. This metric has become a central focus for companies as it directly influences profitability, brand reputation, and customer loyalty. As competition intensifies, businesses increasingly rely on Customer Relationship Management (CRM) systems to collect, analyze, and act upon customer data, which can reveal preferences, buying behaviors, and emerging trends. By capturing these insights, CRM systems allow companies to foster relationships that go beyond transactional interactions, helping them retain loyal customers and strategically position themselves in the market (Zaghloul, Barakat, & Rezk, 2024; Luciano Cavalcante Siebert et al., 2019). With the support of CRM-driven insights, companies can adapt to changing customer needs by delivering personalized services that reinforce customer loyalty and deepen engagement.

The exponential growth of e-commerce, especially accelerated by the COVID-19 pandemic, has led to an unprecedented surge in customer data. This data explosion is primarily driven by online interactions, which produce a vast array of customer insights that can help businesses understand evolving consumer preferences. Traditional data analysis methods, however, struggle to keep pace with the sheer volume and complexity of this information. Machine learning (ML) and deep learning (DL) technologies have emerged as transformative tools in this regard, enabling companies to automate data analysis, detect hidden patterns, and derive predictive insights that guide proactive decision-making. These approaches empower businesses to forecast customer satisfaction levels by identifying key factors that drive satisfaction, such as product quality, customer service responsiveness, and personalization (Liu et al., 2024). In doing so, ML and DL allow companies to enhance the customer experience, making it a fundamental part of strategic growth.

One significant application of machine learning is in analyzing clickstream data—a digital record of customer interactions on websites and applications. Clickstream data provides businesses with a detailed view of the customer journey, from initial browsing to purchase and beyond. By leveraging machine learning on clickstream data, companies can create predictive models that help anticipate customer needs and satisfaction levels. These models not only optimize marketing strategies and tailor product recommendations but also enable companies to address potential issues before they are raised in customer feedback. Consequently, businesses can preemptively improve customer experience and satisfaction, leading to higher retention rates and more favorable brand perceptions (Md Abu Sufian Mozumder et al., 2024). This proactive approach is critical in today’s data-driven world, where companies that fail to leverage customer insights risk losing competitive ground.

The role of predictive analytics in customer satisfaction extends beyond simple analysis, providing actionable insights that inform a company’s overall strategy. For example, predictive models based on machine learning can highlight specific customer pain points, allowing businesses to prioritize improvements in areas such as response times, website usability, or product quality. These insights are invaluable for companies striving to maintain high levels of customer satisfaction and adapt swiftly to market demands. Machine learning tools also enable businesses to segment their customer base more effectively, ensuring that marketing efforts and service improvements target the right groups with the most relevant messaging.

The purpose of this study is to conduct a comparative analysis of machine learning algorithms including Random Forest, Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost) to assess their effectiveness in predicting customer satisfaction. This analysis will focus on identifying the model with the highest accuracy and reliability, which can offer businesses a powerful tool for anticipating customer needs and improving service quality. By evaluating these algorithms, the study aims to provide insights into the most suitable techniques for real-time customer satisfaction prediction. In turn, these findings will empower businesses to harness the full potential of machine learning in fostering customer loyalty, refining service offerings, and maintaining a competitive advantage in a rapidly shifting market. The resulting insights will not only enhance customer experiences but also provide a data-driven foundation for strategic decision-making.

**1.1 Research Question:**

1. Which Machine Learning Algorithm Provides the Most Accurate Prediction of Customer Satisfaction Levels?

2. What Are the Top 5 Key Factors Driving Customer Satisfaction, as identified by Predictive Models

**1.2 Project Objectives**

1.To compare the performance of Random Forest, Support Vector Machine (SVM), and XGBoost in predicting customer satisfaction levels.

2. To evaluate the accuracy, precision, recall, and F1-score of Random Forest, SVM, and XGBoost for customer satisfaction prediction.

3.To identify the top 5 key factors driving customer satisfaction as revealed by Random Forest, SVM, and XGBoost.

4. To determine which of the three algorithms: Random Forest, SVM, or XGBoost provides the most accurate prediction of customer satisfaction levels and recommend the most effective algorithm among them for customer satisfaction prediction.

**Chapter 2: Literature Review**

This chapter presents a review of relevant literature on customer satisfaction analysis, with a particular focus on the evolving role of machine learning (ML) and deep learning (DL) in predicting and interpreting customer feedback. As companies increasingly rely on data-driven insights, traditional methods of measuring customer satisfaction, such as surveys, face challenges of limited response rates and relevance. This review explores emerging ML and DL frameworks that aim to overcome these limitations, offering higher accuracy and greater depth in extracting insights from customer feedback. Through an examination of recent studies, this chapter highlights the strengths and weaknesses of various predictive models, emphasizing their contributions to the broader field of customer satisfaction analysis.

Mozumder et al. (2024) introduces an advanced framework for customer satisfaction analysis that employs Bidirectional Encoder Representations from Transformers (BERT) to classify customer feedback into specific satisfaction drivers. Acknowledging the limitations of traditional customer satisfaction metrics like the American Customer Satisfaction Index (ACSI) and Net Promoter Score (NPS), the authors propose a machine learning-based approach that integrates BERT with term frequency-inverse document frequency (TF-IDF). Using a proprietary dataset of 5,943 customer feedback responses from 39 companies across 13 industries, the study demonstrates BERT's potential to address survey fatigue and increase prediction accuracy.

Results indicate that the fine-tuned BERT model achieved an F1 score of 0.84, outperforming traditional methods such as support vector machine (SVM) and multi-layer perceptron (MLP) networks, which scored 0.47 and 0.50, respectively. A hybrid approach that combined BERT with TF-IDF and MLP reached an F1 score of 0.71, showcasing the effectiveness of hybrid ML approaches. Mozumder et al. conclude that DL techniques like BERT are transformative for CSAT modeling, bridging the gap between traditional and advanced methods while providing actionable insights that support strategic decision-making.

In the highly competitive airline industry, customer satisfaction is essential, as airlines strive to improve services and strengthen relationships with passengers. Elizabeth et al. (2023) conducted a study to predict customer satisfaction based on various parameters, using the Airlines Customer Satisfaction dataset provided by IIT Roorkee in 2020. This dataset includes 129,880 entries and 24 variables, with “satisfaction” as the target. The researchers employed both a Blackbox approach using a deep neural network, which achieved a 92% accuracy, and a Glassbox approach using a decision tree, which reached 94% accuracy. Standard metrics such as accuracy, precision, recall, F1 score, and confusion matrices were used to evaluate the models.

The study further applied Explainable AI (XAI) techniques, including LIME and SHAP, to enhance the interpretability of the Blackbox model. However, since the Glassbox decision tree model is inherently explainable, XAI was not necessary. Given its superior accuracy and interpretability, Elizabeth et al. (2023) recommend the use of the Glassbox approach for predicting airline customer satisfaction. The research emphasizes that explainability in predictive models can be crucial in operational settings, enabling airlines to make actionable improvements based on easily interpretable insights.

In alignment with the focus on customer satisfaction in the airline industry, Amalia, Deborah, and Yulita (2022) underscore the importance of service quality as a driver of customer loyalty. The authors discuss how customer satisfaction data can be leveraged through machine learning models to predict satisfaction levels, thereby guiding management decisions. This study compares the performance of three machine learning models: Split Point and Attribute Reduced Classifier (SPAARC), Multilayer Perceptron (MLP), and Random Forest (RF). Results indicate that the Random Forest algorithm outperforms the other methods, achieving a high accuracy of 95.827%, an F1 score of 0.958, and the shortest training time of 84.53 seconds.

The study emphasizes Random Forest’s balance of accuracy and efficiency, suggesting it as the optimal model for predicting customer satisfaction within airline services. Unlike more computationally intensive models, Random Forest offers a practical solution with minimal training time, which is particularly beneficial for dynamic business environments. This finding supports Elizabeth et al. (2023), who also found decision tree-based models favorable for accuracy and interpretability, further illustrating that explainable machine learning models are advantageous for practical application in customer satisfaction prediction.

The rapid expansion of e-commerce during the COVID-19 pandemic has intensified competition, compelling businesses to prioritize customer satisfaction. Fouad, Barakat, and Rezk (2023) explore methods to enhance e-commerce services by analyzing customer feedback and reviews to predict satisfaction levels. Using five classification algorithms; Decision Tree (DT), Random Forest (RF), XGBoost, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN), the study aims to predict review scores as a proxy for customer satisfaction. After preprocessing the data and engineering new features, the Random Forest model emerged as the most effective, achieving an F1 score of 0.67, outperforming the other algorithms.

The study identifies key factors influencing customer satisfaction in e-commerce, including location, delivery time, product value, and freight ratio, which collectively impact the prediction accuracy. Fouad et al. conclude that the Random Forest model’s performance in predicting customer ratings positions it as a valuable tool for e-commerce platforms. By leveraging this model, businesses can gain actionable insights, allowing them to enhance customer experiences in a competitive online market.

In the highly competitive telecommunications industry, customer satisfaction—particularly concerning call quality is vital for a company's sustainability. Khalid, Jugurnauth, and Muhammad (2022) investigate how machine learning can be leveraged to predict customer satisfaction related to call drop quality across various telecom providers. Using a dataset from Kaggle that includes data samples over three months (September to November), the study applies a Random Forest Classifier to analyze parameters such as call rating, call drop, and the number of subscribers. The model was evaluated through metrics like recall, precision, accuracy, and F1-score, achieving an accuracy of 91% in predicting customer satisfaction regarding call quality.

Unlike studies that compare multiple algorithms, this research solely focuses on the Random Forest Classifier, underscoring its efficacy in assessing telecom service quality. Khalid et al. argue that this model can offer telecom providers actionable insights into call quality factors affecting customer retention and satisfaction. The study demonstrates how predictive models can help telecom companies adapt to customer demands, minimizing service quality issues that might otherwise lead to customer churn or market exit.

The recent entry of Air Peace airline into the London market has intensified competition, emphasizing the importance of customer satisfaction in gaining a competitive edge. Sobowale et al. (2024) explore how machine learning can improve customer satisfaction prediction within the airline industry, using Support Vector Machine (SVM) classifiers to analyze customer feedback on hospitality services. An open-access dataset, containing attributes such as age, gender, and flight distance, was pre-processed through techniques like normalization, label encoding, and feature selection. The study compares SVM with three kernel types Radial Basis Function (RBF), Linear, and Polynomial using an 80-20 train-test split for hold-out evaluation, achieving average accuracies of 93%, 85%, and 90%, respectively.

The findings show that the RBF kernel performed best, with an accuracy of 93% alongside precision and F1 scores of 93% and 92%, respectively. Sobowale et al. conclude that SVM with the RBF kernel provides accurate and reliable insights into customer satisfaction factors, enabling airlines to identify and prioritize key service elements for enhanced customer experience. The research underscores how machine learning models can optimize customer service strategies, improving resource allocation and overall service delivery during peak periods.

Airline customer satisfaction, encompassing factors like booking processes, in-flight experience, and customer service, is essential for building brand loyalty and sustaining an airline's reputation. In a competitive aviation market, high levels of customer satisfaction are critical for long-term viability. Kumar et al. (2024) investigate this by analyzing a large dataset of 129,880 customer records using machine learning models, including logistic regression, decision tree, and random forest. The study reveals that the random forest model outperformed the other methods, achieving an accuracy exceeding 94%, thus underscoring its robustness in predicting customer satisfaction. By focusing on customer satisfaction metrics and various service aspects, the study highlights the value of advanced machine learning approaches in the airline industry. Kumar et al. conclude that machine learning, particularly the random forest model, offers precise and reliable predictions, making it a valuable tool for airlines aiming to improve service quality and foster customer loyalty. This research aligns with previous studies on the efficacy of machine learning in enhancing decision-making based on customer feedback, solidifying random forest's role as an optimal choice for predictive analysis in customer satisfaction.

Understanding the underlying factors of customer satisfaction is essential for both academics and service industry practitioners. Aldunate et al. (2022) introduce an innovative framework that leverages deep learning models to automatically analyze open-ended survey data and extract key drivers of customer satisfaction. Based on 11 drivers identified in marketing literature as influential to customer experience, this framework employs a multi-label classification approach, which enables it to analyze and classify responses from diverse service industries. Using 25,943 customer survey responses from 39 companies across 13 economic sectors, the model demonstrates high accuracy in identifying satisfaction drivers. This deep learning-based framework provides not only automated analysis but also ranks the importance of each driver across different service sectors, offering valuable insights for targeted customer experience improvements. Aldunate et al. conclude that such models can streamline data processing, giving businesses a deeper understanding of what factors impact customer satisfaction the most. This research emphasizes the potential of natural language processing in simplifying and enhancing customer feedback analysis, supporting strategic decisions in customer satisfaction management.

Recognizing that customer satisfaction prediction is central to business success, Xiao et al. (2022) conducts a comprehensive review of machine learning methods applied across various customer satisfaction studies. They emphasize that different data formats, such as review and survey data, necessitate the selection of distinct machine learning models to optimize prediction accuracy. The study demonstrates that customer satisfaction prediction can be effectively analyzed using a range of typical algorithms, each tailored to the data type and specific application field. Xiao et al. conclude that model selection is crucial for accurate customer satisfaction analysis, with different algorithms recommended for different data structures and industries. This research provides guidance for businesses seeking to implement machine learning models for customer satisfaction prediction, underscoring the importance of aligning model choice with data format to achieve the best outcomes. By highlighting the nuances of data-driven model selection, the study contributes to a more refined approach to customer satisfaction analytics.

The integration of AI tools in online shopping has demonstrated a significant positive impact on customer satisfaction, particularly through the application of machine learning techniques. Kannan et al. (2023) explore how online retailers can leverage AI-driven insights to predict satisfied shoppers with high accuracy. This study collected data from online shoppers in India, China, and Canada through surveys conducted between 2021 and 2022, focusing on the shoppers’ experiences with AI tools. Five machine learning algorithms were tested: decision tree, random forest, naïve Bayes, gradient-boosted tree, and multilayer perceptron neural network. Among these, the random forest model achieved the highest performance, with an F1-score of 91.5%, while all models exceeded an accuracy threshold of 86.5%.

The findings suggest that machine learning enables online retailers to identify satisfied customers with a high degree of accuracy, which can drive data-informed improvements in customer experience. Kannan et al. argue that integrating machine learning into online retail not only boosts customer satisfaction but also enhances business efficiency and competitiveness. This study highlights the transformative potential of AI tools in fostering more tailored and innovative customer care within the online shopping landscape.

**References**

Amalia, S., Deborah, I. and Yulita, I.N. (2022). Comparative analysis of classification algorithm: Random Forest, SPAARC, and MLP for airlines customer satisfaction. *SINERGI*, 26(2), p.213. doi:https://doi.org/10.22441/sinergi.2022.2.010.

Aldunate, Á., Maldonado, S., Vairetti, C. and Armelini, G. (2022). Understanding customer satisfaction via deep learning and natural language processing. *Expert Systems with Applications*, 209, p.118309. doi:https://doi.org/10.1016/j.eswa.2022.118309.

‌

Elizabeth, Yong Wan Yit, Zaur Tahirovich Gouliev and Quille, K. (2023). Predictive Models with XAI: A Comparative Study of Enhancing Airline Customer Satisfaction. doi:https://doi.org/10.1145/3633083.3633189.

Fouad, M., Barakat, S. and Rezk, A. (2023). Effective E-commerce Based on Predicting the Level of Consumer Satisfaction. *Lecture notes in networks and systems*, pp.261–278. doi:https://doi.org/10.1007/978-981-99-4764-5\_17.

‌ Zaghloul, M., Barakat, S. and Rezk, A. (2024). Predicting

Liu, D., Huang, H., Zhang, H., Luo, X. and Fan, Z. (2024). Enhancing customer behavior prediction in e-commerce: A comparative analysis of machine learning and deep learning models. *Applied and Computational Engineering*, 55(1), pp.190–204. doi:https://doi.org/10.54254/2755-2721/55/20241475.

‌Luciano Cavalcante Siebert, Bianchi, F., Júnior, S., Eduardo Kazumi Yamakawa and Catapan, A. (2019). Predicting customer satisfaction for distribution companies using machine learning. *International Journal of Energy Sector Management*, 15(4), pp.743–764. doi:https://doi.org/10.1108/ijesm-10-2018-0007.

‌Md Abu Sufian Mozumder, Tuan Ngoc Nguyen, Suniti Devi, Md Arif, Md Parvez Ahmed, Estak Ahmed, Maniruzzaman Bhuiyan, Md Habibur Rahman, Abdullah Al Mamun and Aftab Uddin (2024). Enhancing Customer Satisfaction Analysis Using Advanced Machine Learning Techniques in Fintech Industry. *Journal of Computer Science and Technology Studies*, 6(3), pp.35–41. doi:https://doi.org/10.32996/jcsts.2024.6.3.4.

Kannan, Rathimala; Ramakrishnan, Kannan; Ersoy, Ayse Begum; Contu, Davide (2023). Customer Satisfaction Response to Artificial Intelligence Tools Usage During Online Shopping. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.24633105.v1>

Khalid, Jugurnauth, R.A. and Muhammad (2022). Application of Machine Learning in Predicting Customer Satisfaction of Telecom Service Providers. doi:https://doi.org/10.1109/elecom54934.2022.9965212.

Kumar, A., Mehta, V., Rao, T. and Ashwini Kodipalli (2024). Forecasting Airline Customer Satisfaction through the Application of Diverse Machine Learning Algorithms. doi:https://doi.org/10.1109/icetcs61022.2024.10543367.

‌Sobowale, A.A., Osadare, O.O., Soladoye, A.A. and Sobowale, P.O. (2024). Prediction of Customer Satisfaction in Airline Hospitality services for improved service delivery using Support Vector Machine. *FUOYE Journal of Engineering and Technology*, 9(2), pp.241–245. doi:https://doi.org/10.4314/fuoyejet.v9i2.13.

‌Xiao, G., Kuan, T.-W., Li, F., Ling, M. and Tseng, S.-P. (2022). *Systematic Review of Machine Learning Methods in Customer Satisfaction Prediction*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICOT56925.2022.10008132.

‌Zaghloul, M., Barakat, S. and Rezk, A. (2024). Predicting E-commerce customer satisfaction: Traditional machine learning vs. deep learning approaches. *Journal of Retailing and Consumer Services*, [online] 79, p.103865. doi:https://doi.org/10.1016/j.jretconser.2024.103865.

‌