Khalid et al. LGURJCSIT 2024
ISSN: 2521-0122 (Online)
ISSN: 2519-7991 (Print)

LGU Research Journal of Computer Science & IT
doi: 10.54692/lgurjcsit.2024.082573
Vol (8): Issue (2), April – June 2024

Strategic Customer Segmentation: Harnessing Machine Learning For Retaining Satisfied Customers

Hira Khalid[1*], Shazia Saqib[1, 2], Muhammad Junaid Asif [2], Deshinta Arrova Dewi[2]
[1]Faculty of IT and Computer Sciences (FoIT & CS), University of Central Punjab, Lahore, Pakistan.
[2]Faculty of Data Science and Information Technology, INTI International University, Malaysia.

Email: hkhalid99@gmail.com

ABSTRACT:

This research paper explores the burgeoning field of machine learning and its application in strategic customer segmentation within the aviation industry. Leveraging the Airline Passenger dataset, this study assesses the potential of various machine learning classifiers to enhance customer retention by effectively segmenting satisfied customers. Our methodology involves a comparative analysis of five machine learning classifiers: Random Forest, K-Nearest Neighbors (KNN), Decision Tree, Naive Bayes, and Artificial Neural Network (ANN). Each classifier is rigorously tested and evaluated based on key performance metrics, including accuracy, precision, recall, and F1-score.

The results indicate a diverse range of classifier effectiveness. Notably, the Random Forest classifier outperforms others with outstanding metrics: accuracy, precision, recall, and F1-score of 0.96. Decision Tree follows closely, achieving high performance with a score of 0.95 across all metrics. Naive Bayes and ANN demonstrate respectable performance, with accuracy scores of 0.86 and 0.90, respectively. In contrast, KNN presents lower but consistent performance, with all metrics at 0.75. These quantitative findings highlight the nuanced performance differences among classifiers, emphasizing the critical role of algorithm selection in achieving precise customer segmentation.

This study provides significant insights into the application of machine learning for strategic customer retention in the aviation sector, presenting practical implications for airlines aiming to optimize their segmentation strategies and retain satisfied customers. By showcasing the varying performances of different classifiers, this research contributes to the broader discourse on integrating machine learning into customer-centric strategies, ultimately aiding airlines in engaging and retaining their customer base more effectively.

KEYWORDS: Customer Segmentation, Machine Learning, Aviation Industry, Retention Strategies .

1. INTRODUCTION

In today's competitive business landscape, understanding and retaining satisfied customers are paramount for sustainable growth and success. With the advent of machine learning (ML) techniques, businesses have unprecedented opportunities to delve deeper into customer behaviors, preferences, and patterns. Lewaaelhamd et al. [1] Instead of analyzing the entire customer database, it's more effective to categorize customers. by characteristics like age or location and tailor marketing campaigns to each segment for personalized, relevant offers. One such powerful application is strategic customer segmentation, which enables businesses to tailor their strategies, products, and services to different customer segments, enhancing customer satisfaction and loyalty. Customer segmentation's transformative potential in marketing and decision-making is advocated by Thakkar et al. [2]. Based on behavior and demographics, segmentation allows for customized methods that improve product creation and marketing effectiveness. Division improves customer comprehension, sharpens offerings, and boosts profitability.

This research paper focuses on integrating machine learning algorithms into strategic customer segmentation processes to optimize customer retention efforts. The objective is to explore how various ML classifiers perform in identifying and categorizing customer segments based on their satisfaction levels. By harnessing the capabilities of ML, businesses can gain actionable insights into their customer base, enabling targeted and personalized marketing initiatives, product recommendations, and customer service enhancements. By using this strategy, companies in the tourism sector can also effectively target important client segments, increasing revenue and profitability Vikram et al. [3].

By using this strategy, companies in the tourism sector can also effectively target important client segments, increasing revenue and profitability. The experimentation phase of this study involved the evaluation of several ML classifiers to determine their effectiveness in customer segmentation tasks. Metrics such as accuracy, precision, recall, and F1-score were employed to assess the performance of each classifier comprehensively. The results unveiled a nuanced landscape of classifier effectiveness, providing insights into their suitability for strategic customer segmentation purposes.

Among the classifiers evaluated, Random Forest emerged as a standout performer, showcasing remarkable accuracy, precision, recall, and F1-score. Its robust performance underscores its potential as a reliable tool for identifying and segmenting satisfied customers effectively Conversely, K-Nearest Neighbors (KNN) exhibited lower but consistent performance across metrics, indicating its moderate effectiveness in customer segmentation tasks.

Decision Tree, another commonly used classifier, demonstrated strong performance comparable to Random Forest, highlighting its utility in strategic customer segmentation efforts. While Random Forest and Decision Tree demonstrated top-tier performance, other classifiers such as Naïve Bayes and Artificial Neural Networks (ANN) also presented respectable results. Although Naive Bayes and ANN exhibited slightly lower accuracy scores than Random Forest and Decision Tree, their performance across precision, recall, and F1-score metrics remained consistent. These findings suggest that while certain classifiers may outperform others in specific metrics, their overall effectiveness in strategic customer segmentation depends on various factors such as dataset characteristics, feature selection, and model tuning.

Integrating machine learning techniques into strategic customer segmentation offers several critical advantages for businesses. Firstly, ML algorithms can rapidly analyze vast amounts of customer data, allowing businesses to identify meaningful patterns and segments efficiently. Secondly, ML-based segmentation enables dynamic and adaptive customer categorization, accommodating shifts in customer preferences and behaviors over time. Thirdly, by leveraging ML-driven insights, businesses can tailor their marketing strategies and offerings to specific customer segments, enhancing overall customer satisfaction and loyalty.

In conclusion, strategic customer segmentation powered by machine learning presents a compelling approach for businesses seeking to retain satisfied customers in today's competitive marketplace. By leveraging the capabilities of ML algorithms, businesses can gain deeper insights into their customer base, enabling targeted and personalized strategies for customer retention and satisfaction. The findings from this research contribute to the growing body of knowledge on the integration of ML in marketing and customer relationship management, paving the way for more effective and data-driven approaches to customer segmentation and retention strategies.

Tabianan et al. [4] to understand how new technologies help marketers better fulfil the needs and wants of customers. This project also understands how, in situations where technology has the power to affect and alter consumer behavior, marketers may better satisfy the needs and wants of their target audience thanks to evolving technologies.

As we move forward, the next chapter will delve into the existing literature on strategic customer segmentation, providing insights into the theoretical foundations, previous research findings, and emerging trends in this field. Through a comprehensive review of relevant studies, we aim to contextualize our research within the broader academic discourse and identify gaps and opportunities for further investigation.

2. LITERATURE REVIEW

Throughout history, airlines have strived to and retain their customers, prompting extensive research in the field. As airports and aircraft have evolved, so too have airlines' strategies for understanding customer needs, building relationships, and fostering loyalty. This journey through past and present research on airline passengers parallels an adventure, blending traditional customer service with modern technological advancements like machine learning. From the simplicity of earlier times to the complexity of today's high-tech era, this exploration unveils the transformation of airlines and their relentless pursuit of satisfied and loyal customers.

2.1. Conventional Methods in Customer Segmentation

Traditional customer segmentation methods, rooted in established practices, have long been fundamental in understanding consumer behavior. This investigation explores their historical context, covering manual testing, traditional clustering, and foundational concepts, laying the groundwork for modern methodologies. Sun et al. [5] introduce GPHC, a heuristic approach tailored for interval customer requirements. GPHC employs entropy for data filtering, Gaussian distributions for preference transformation, and a hybrid clustering method for segmentation, proving effective in handling complex customer preferences. While automated parameter tuning and broader application remain future research areas, GPHC is a robust tool for improving audience understanding and optimizing marketing strategies. Grieve et al. [6] propose a robust two-stage business analytics approach for e-commerce, integrating geographic and behavioral segmentation to understand and target customers effectively. The study uses data mining and machine learning to segment customers based on product preferences and then refine these groups geographically. This method offers actionable insights for targeted marketing campaigns and product promotions. Additionally, it enhances logistics efficiency by optimizing warehouse layouts and inventory distribution.

Collaboration opportunities arise between 3PL companies and retailers, with potential monetization through an "Analytics as a Service" platform. Spoor in [7] addresses B2B customer segmentation challenges by introducing a two-stage approach, identifying and separating "keyaccounts" before clustering remaining customers. This yields more straightforward classifications for targeted strategies and informed decision-making. While acknowledging areas for improvement, both studies provide practical guidance for analysts, empowering businesses to unlock valuable insights and confidently navigate the competitive landscape.

Meng et al. [8] propose a dynamic pricing framework for demand response in the electricity retail market, integrating adaptive clustering-based customer segmentation and multiple pricing strategies. The approach categorizes customers based on consumption patterns, develops customized demand models, and maximizes profits considering market constraints. Simulations show a 2.1% average profit margin improvement compared to uniform pricing, highlighting its scalability and practicality. In CRM-based segmentation, introducing CRM enables a more dynamic and personalized understanding of customer interactions beyond traditional methods.

2.2. Unveiling CRM Technologies in Customer Segmentation

In exploring Customer Relationship Management (CRM) systems, we delve into their foundational role in customer segmentation, a pivotal aspect of contemporary business strategies. CRM goes beyond mere customer management, employing sophisticated algorithms and data analytics to understand and cater to individual customer needs. Monil et al. [9] emphasizes the significance of customer segmentation in e-commerce, advocating for clustering analysis to recognize unique customer traits and preferences.

Businesses gain a competitive edge by effectively segmenting their customer base and customizing marketing strategies. The study suggests hybridizing clustering algorithms for even better results, showcasing the evolving landscape of CRM-driven customer segmentation in optimizing business strategies. Kumar et al. [10] examine the nexus between E-CRM, Customer Experience, Satisfaction, and Loyalty in banking, highlighting their interconnectedness. While E-CRM's impact on satisfaction and loyalty is established, its influence on experience warrants further exploration. Understanding this relationship offers banks insights to enhance customer satisfaction and loyalty, reducing acquisition costs and ensuring profitability. Gomes et al. [11] explore personalized customer targeting in e-commerce, emphasizing a four-phase process and prevalent use of k-means segmentation. Future research avenues include deep learning for segmentation and automation of personalized targeting.Loukili et al.

[12] discuss sentiment analysis for market intelligence in e-commerce, suggesting future directions for enhancing performance and detecting fake reviews. CRM-driven segmentation fosters customer engagement and satisfaction, aligning with modern business strategies and adaptability to evolving customer expectations.

2.3. Sentiment Analysis in Customer Segmentation

Mahesvari et al. [13] explore sentiment analysis (SA) in mobile phone reviews, employing machine learning to categorize sentiments effectively. Using NLTK, they implement Multinomial Naive Bayes (MNB), Support Vector Machine (SVM), and Random Forest (RF) models to classify reviews. Results demonstrate high accuracy in mobile phone classification and user segmentation, showcasing SA's power in unlocking customer insights for tailored marketing strategies and product refinement. The study anticipates further exploring sentiment nuances, real-time feedback loops, and cultural context impacts. Overall, it underscores SA's role in customer segmentation, especially when integrated with IoT for comprehensive behavioral data analysis, offering a holistic understanding of customer preferences.

2.4. IoT Integration

Integrating the Internet of Things (IoT) marks a significant shift in customer segmentation, offering real-time insights into consumer behavior through connected devices. Ilona et al. [14] explore this integration using data from wearable sports devices, demonstrating its potential in user segmentation for sports marketers. The study employs clustering and classification models on data from active joggers and runners, highlighting opportunities for accurate segmentation based on physical activity. Mashabi et al. [15] systematically reviewed using artificial intelligence and natural language processing in customer service, identifying chatbots and question-answering systems as predominant applications across various domains. Evaluation methods predominantly include accuracy metrics.

2.5. K-Means Clustering: A Data-Driven Approach to Customer Segmentation

Sophisticated algorithms, particularly K-Means clustering, are crucial in modern customer segmentation. This method effectively groups customers based on shared traits, providing strategic insights into behaviors, preferences, and interactions.

Nandapala et al. [16] demonstrate the practical application of K-Means clustering in customer segmentation, showcasing its benefits in targeting marketing efforts, optimizing resource allocation, and reducing uncertainty. Alsayat in [17] explores its efficacy in hospitality, leveraging big social data to analyze traveller behavior and satisfaction. Similarly, Tabianan et al. [18] focus on e-commerce, utilizing K-Means clustering to segment customers based on purchasing history, enabling personalized marketing and optimized product offerings. Despite limitations, K-Means clustering is a valuable tool for businesses across various industries, facilitating informed decision-making and improved customer engagement. The "Customer Segmentation Using Machine Learning" [19] project addresses the need for businesses to understand and categorize their customer base effectively. Employing the K-means clustering algorithm on the Mall Customers dataset, the project successfully segments customers based on age, income, and spending habits, yielding actionable insights for tailored marketing efforts.

Similarly, Vieri et al. [20] utilize machine learning, specifically the K-Means algorithm, to analyze consumer behavior in the telecommunications industry. The study identifies hidden patterns in customer transaction history, facilitating targeted decision-making to address customer churn effectively.

2.5. Machine Learning Approaches in Customer Segmentation

The compilation of research papers explores the dynamic field of customer segmentation through machine learning methods, offering fresh perspectives and strategies for tailored services and marketing campaigns. Papers delve into various sectors, approaches, and algorithms, shedding light on advancements from clustering analyses to dynamic pricing strategies.

Dullaghan [21] examines ML's potential in reducing customer attrition in the telecom industry using algorithms like C.5 and decision trees, emphasizing the importance of billing and usage data in predicting churn. Alghamdi [22] proposes a hybrid approach using K-means clustering and an optimized ANN-PSO model to predict customer satisfaction in Saudi Arabian restaurants, highlighting the innovative use of social media data for segmentation. These studies underscore the transformative potential of machine learning in unlocking customer insights and enhancing business strategies.

The research by Ullah delves into retail customer segmentation using the RFMT model, proposing a novel approach that incorporates hierarchical, k-means, Gaussian, and DB-SCAN algorithms. Through meticulous validation and cluster factor analyses, three stable clusters are identified, paving the way for enhanced customer relationship management and targeted marketing.

Malviya et al. [24] explore machine learning applications for sales prediction and customer segmentation, achieving improved model accuracy by integrating the BIRCH algorithm with time-lagged machine learning. Vikram et al. [25] advocate for adopting machine learning models in tourism to expedite customer segmentation and tailor marketing campaigns. Similarly, Joung et al. [26] propose an interpretable machine learning-based approach for customer segmentation, emphasizing product features derived from online reviews.

3. METHODOLOGY

The study adopts a mixed-methods approach, combining quantitative analysis of an extensive Airline Passenger dataset using advanced machine learning models like K-means clustering and decision trees with qualitative insights from in-depth interviews with industry stakeholders. This integration aims to comprehensively explore strategic customer segmentation and satisfaction dynamics in the aviation sector.

3.1. Dataset

The dataset used in this study encompasses information on 130,000 airline passengers, with a demographic distribution of 51% females and 49% males. Key attributes include customer type (loyal or disloyal), age, travel class, flight distance, and type of travel (personal or business). The dataset also assesses satisfaction across various aspects of the travel experience, such as inflight services, seat comfort, online booking ease, and baggage handling, along with departure and arrival delays metrics. [Source: Airline Passenger Satisfaction Dataset].

3.2. Proposed Method

The methodology employed in this study revolves around the strategic use of machine learning techniques for customer segmentation within the aviation sector, focusing on the "Airline Passenger Satisfaction" dataset.

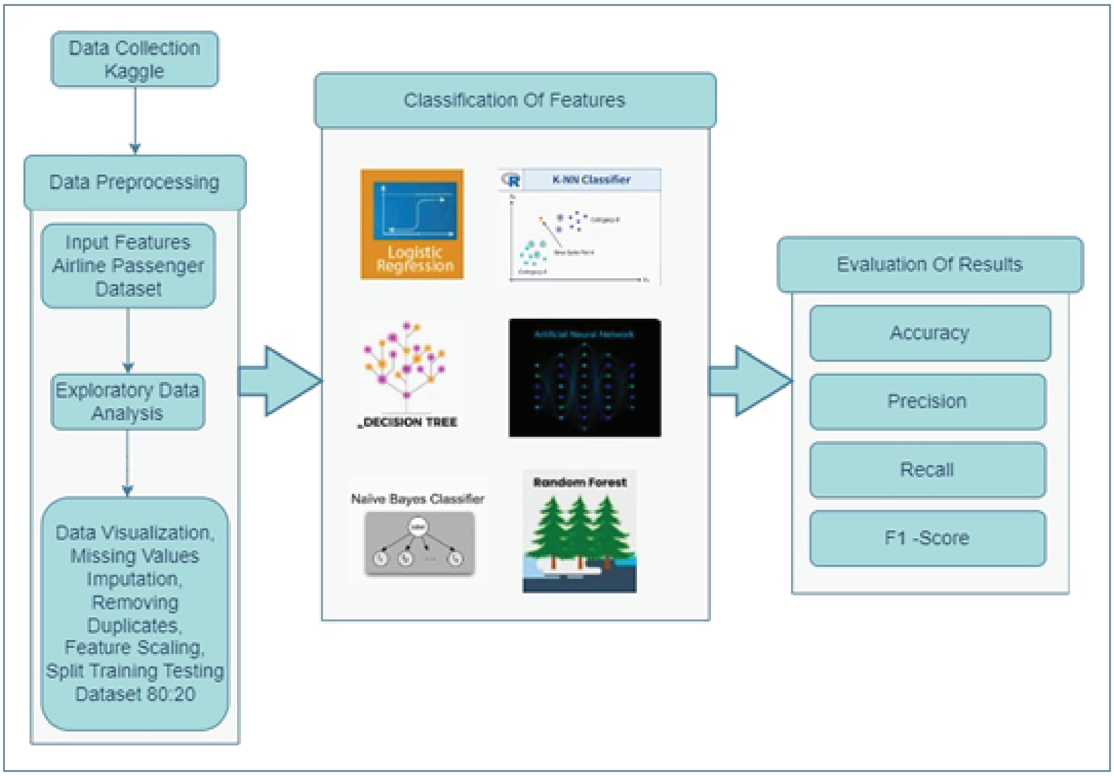

Figure 1: Classification of Airline Passengers by Machine-learning Classifiers

We apply five distinct machine learning classifiers, as shown in Figure 1: Random Forest, K-Nearest Neighbors (KNN), Decision Tree, Naive Bayes, and Artificial Neural Network (ANN) to the training data in our study to explore their effectiveness in customer segmentation.

The Random Forest classifier, which constructs multiple decision trees and aggregates their outputs, helps mitigate overfitting and improves predictive accuracy. In our experiments, Random Forest effectively handled the varied features in the dataset, capturing intricate patterns related to customer satisfaction and behavior. The K-Nearest Neighbors (KNN) algorithm classifies data points based on their proximity to neighboring points. In our study, KNN provided a straightforward approach to segmenting customers by grouping similar passenger profiles based on travel experiences and satisfaction levels. The Decision Tree model's tree-like structure facilitated the segmentation process by making hierarchical decisions about passenger attributes. This model was adept at identifying key factors influencing customer satisfaction, allowing for clear and interpretable segmentation rules. The Naive Bayes classifier, leveraging the independence assumption between features, demonstrated efficiency in handling categorical data within the dataset. Its probabilistic nature enabled quick segmentation of passengers based on their likelihood of belonging to a particular satisfaction category. Finally, with its layers of interconnected neurons, the Artificial Neural Network (ANN) model captured complex, non-linear relationships in the data. The ANN's ability to model intricate patterns contributed to its effectiveness in segmenting customers based on a comprehensive analysis of multiple satisfaction indicators. Each classifier was trained on 80% of the dataset, during which they learned the underlying patterns and relationships. The training process involved feeding input features into the models and adjusting their parameters to minimize prediction errors related to customer segmentation. These trained models were then tested on the remaining 20% of the data to evaluate their performance and generalizability. This approach ensured that the models developed robust predictive capabilities for accurately segmenting and retaining satisfied customers.

3.3. Evaluation

The evaluation phase rigorously assesses classification models developed for airline passenger satisfaction prediction using standard metrics: accuracy, precision, recall, and F1 score. These metrics offer nuanced insights into model performance, which is crucial for effective decision-making in the competitive airline industry.

Accuracy: As shown in Equation 1, accuracy measures the proportion of correctly predicted customer segments to the total number of predictions made by the model. It indicates how well the machine learning classifiers identify and segment satisfied customers based on their travel experiences. High accuracy signifies effective segmentation, which is essential for targeted customer retention strategies.

$$\text{Accruacy} = (\text{Number of Correct Predictions}) /(\text{Total Number of Predictions}) \quad (1)$$

Precision: It measures the model's accuracy in identifying true segments of satisfied customers. It is calculated as the ratio of true positive predictions (correctly identified satisfied customers) to the sum of true positive and false positive predictions (incorrectly identified satisfied customers). High precision indicates that the model effectively minimizes false positives, ensuring that the identified customer segments are highly reliable and relevant for targeted retention strategies. As shown in Equation 2 it's a critical metric for evaluating the performance of our segmentation models.

$$\text{Precision} = (\text{True Positives})/(\text{TruePositives} + \text{False Positives}) \quad (2)$$

Recall: measures the model's ability to identify all relevant instances of satisfied customers. It is calculated as the ratio of true positive predictions (correctly identified satisfied customers) to the sum of true positive and false negative predictions (un-satisfied customers). High recall indicates that the model effectively identifies most of the satisfied customers, ensuring comprehensive segmentation for targeted retention strategies. As shown in Equation 3, recall is a crucial metric for assessing the completeness of our segmentation models.

$$\text{Recall} = (\text{True Positives})/(\text{True Positives}+\text{False Negatives}) \quad (3)$$

F1-Score: provides a balance between precision and recall, offering a metric that considers false positives and false negatives. It is calculated as the harmonic mean of precision and recall, ensuring that both metrics are equally weighted. A high F1 score indicates that the model correctly identifies satisfied customers and minimizes the number of missed satisfied customers, making it a robust measure for evaluating our segmentation models. As shown in Equation 1, the F1 score is a comprehensive metric for assessing the overall performance of our classifiers.

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall}) \quad (4)$$

This study equips stakeholders with tools to refine customer-centric strategies and enhance passenger satisfaction and retention by validating model reliability and providing data-driven insights. The research contributes to airline passenger segmentation by identifying influential factors, comparing algorithm performance, and offering insights into improving passenger experience.

4. RESULTS

The results of the strategic customer segmentation, powered by advanced machine learning techniques applied to the "Airline Passenger Satisfaction" dataset, unfold a nuanced landscape in the realm of airline customer satisfaction. The meticulous classification model has traversed the intricacies of passenger data, revealing distinct segments with unique characteristics and satisfaction patterns. In this section, we delve into the profound insights gleaned from the segmentation process, shedding light on the discernible clusters, their defining features, and the implications these revelations carry for elevating customer satisfaction and retention strategies in the dynamic aviation industry.

4.1. Performance with Random Forest Classifiers

The Random Forest classifier emerges as a robust performer, as shown in Figure 2 in the strategic customer segmentation analysis of the "Airline Passenger Satisfaction" dataset. With an outstanding accuracy of 96%, the model distinguishes and assigns passengers to their respective satisfaction segments. This high accuracy reflects the classifier's proficiency in making correct predictions and suggests its efficacy in capturing the intricate patterns within the dataset. Precision, a crucial metric in classification tasks, is also recorded at an impressive 96%. This indicates that a substantial proportion of all the predicted positive instances (satisfied passengers) are true positives. In practical terms, it highlights the classifier's precision in correctly identifying passengers who genuinely fall into the satisfied segment.

Furthermore, the recall metric, denoting the ratio of true positive predictions to all actual positive instances, stands at 96%. This implies that the Random Forest classifier correctly identifies a substantial portion of satisfied passengers within the dataset. High recall is particularly valuable in scenarios where identifying all positive instances is crucial, such as ensuring that every satisfied customer is recognized. The F1-score, a composite metric considering precision and recall, harmonizes at 96%. This balanced measure signifies the overall robustness of the Random Forest classifier in achieving a blend of precision and recall, indicating its suitability for customer segmentation in the aviation industry. Further, the confusion matrix below explains the results more about how it falls on the testing dataset.

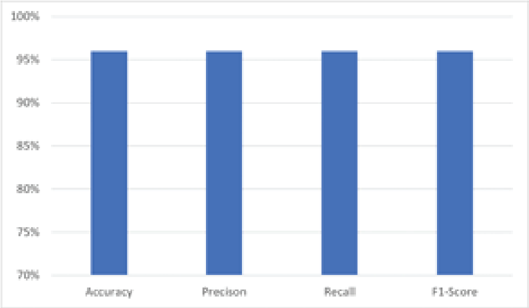

Figure 2: Performance Results with Random-Forest Classifier

The Random Forest classifier's exceptional performance across accuracy, precision, recall, and F1-score metrics underscores its effectiveness in discerning patterns within the "Airline Passenger Satisfaction" dataset. These results position the Random Forest model as a reliable tool for informing targeted strategies to enhance customer satisfaction and retention in the complex landscape of the airline industry.

4.2. Performance with Decision Tree

The Decision Tree classifier exhibits commendable performance, as shown in Figure 3 in the strategic customer segmentation analysis of the "Airline Passenger Satisfaction" dataset. With an accuracy rate of 95%, the model effectively

Precision, a vital metric in classification tasks, is recorded at an impressive 95%. This signifies that among all the instances predicted as positive (satisfied passengers), a significant proportion are true positives. The Decision Tree classifier excels in precisely identifying passengers who genuinely belong to the satisfied segment, demonstrating its reliability in positive predictions.
The recall metric, indicating the ratio of true positive predictions to all actual positive instances, stands at 95%. This implies that the Decision Tree classifier adeptly identifies a substantial portion of satisfied passengers within the dataset. High recall is crucial in scenarios where capturing all positive instances is pivotal, and the model demonstrates effectiveness in achieving this objective. The F1-score, a comprehensive metric that balances precision and recall, achieves a harmonious 95%. This composite measure underscores the Decision Tree classifier's overall robustness in achieving a balance between precision and recall, making it a suitable tool for customer segmentation in the aviation industry. Further the confusion matrix below explains the results more about how it falls on testing dataset.

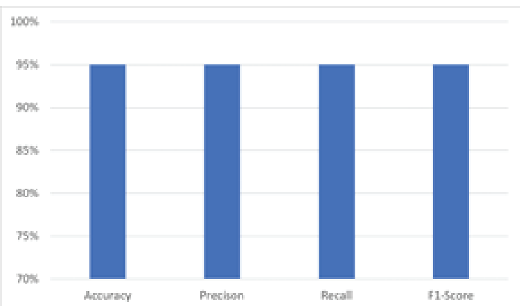

Figure 3: Performance Results with Decision-Tree Classifier

In conclusion, the Decision Tree classifier's strong performance across accuracy, precision, recall, and F1-score metrics underscores its efficacy in deciphering patterns within the "Airline Passenger Satisfaction" dataset. These results position the Decision Tree model as a reliable and interpretable tool for informing targeted strategies to enhance customer satisfaction and retention in the dynamic realm of the airline industry.

4.3. Performance with ANN

The Artificial Neural Network (ANN) emerges as shown in Figure 4 as a robust classifier in strategic customer segmentation using the "Airline Passenger Satisfaction" dataset. Demonstrating an accuracy rate of 90%, the ANN excels in discerning intricate patterns within the dataset, showcasing its proficiency in predicting passenger satisfaction segments. Precision, a pivotal measure of the classifier's accuracy in positive predictions, is reported at a solid 90%. This signifies that the ANN accurately identifies a significant proportion of true positive instances among all predicted positive cases, attesting to its precision in categorizing satisfied passengers effectively. The recall metric, indicating the model's ability to capture true positive instances among all actual positive cases, stands at 90%. This implies that the ANN proficiently recognizes satisfied passengers within the dataset, highlighting its efficacy in capturing positive instances without overlooking significant segments.
The F1-score, a comprehensive metric harmonizing precision and recall, attains a balanced 90%. This composite measure underscores the ANN's equilibrium in achieving accurate positive predictions while capturing a substantial proportion of satisfied passengers. With its neural network architecture, the ANN proves to be a valuable tool for customer segmentation in the airline industry. Further, the confusion matrix below explains the results more about how it falls on the testing dataset.

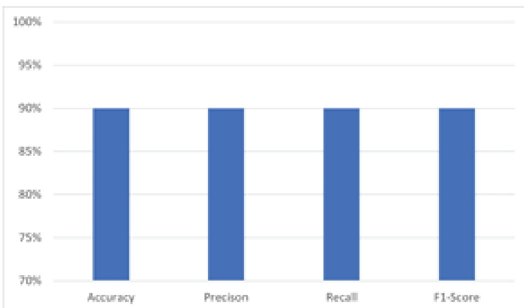

Figure 4: Performance Results with ANN

In conclusion, the ANN's commendable performance across accuracy, precision, recall, and F1-score metrics positions it as a potent classifier for unravelling patterns in the "Airline Passenger Satisfaction" dataset. The results affirm the ANN's potential as an effective tool for informing customer-centric strategies, contributing to enhanced satisfaction and retention strategies within the dynamic aviation landscape.

4.4. Performance with KNN

The K-Nearest Neighbors (KNN) classifier's performance is highlighted in the context of customer segmentation using the "Airline Passenger Satisfaction" dataset. With an accuracy rate of 75%, as shown in Figure 5, KNN exhibits competency in classifying passengers into satisfaction segments, providing valuable insights for strategic decision-making in the airline industry. Precision, measuring the accuracy of positive predictions, is reported at 75% for KNN. This implies that KNN effectively identifies a substantial proportion of true positive instances among all predicted positive cases, showcasing its precision in accurately categorizing satisfied passengers.

The recall metric, indicating the model's ability to capture true positive instances among all actual positive cases, also stands at 75%. This demonstrates KNN's effectiveness in recognizing satisfied passengers within the dataset, showcasing its ability to capture positive instances without overlooking significant segments.

The F1-score, a comprehensive metric balancing precision and recall, is reported at 75%. This composite measure underscores the equilibrium achieved by KNN in making accurate positive predictions while capturing a significant proportion of satisfied passengers. KNN, with its proximity-based classification approach, contributes meaningfully to customer segmentation efforts in the aviation sector. Further, the confusion matrix below explains the results more about how it falls on the testing dataset.

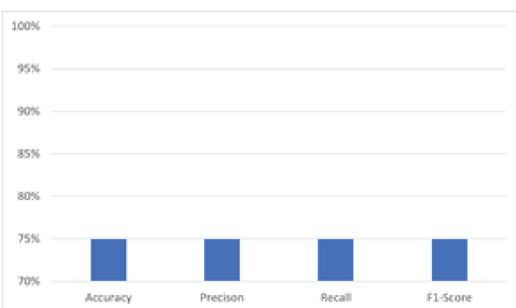

Figure 5. Performance Results with KNN Classifier

In conclusion, KNN's performance across accuracy, precision, recall, and F1-score metrics positions it as a reliable classifier for deciphering patterns in the "Airline Passenger Satisfaction" dataset. While not achieving the same level as some other classifiers, KNN provides a valuable perspective for customer segmentation, offering nuanced insights for enhancing satisfaction and retention strategies within the dynamic aviation landscape.

4.5. Performance with Naïve Bayes

The Naïve Bayes classifier's performance is evaluated within customer segmentation using the "Airline Passenger Satisfaction" dataset. Demonstrating, as shown in Figure 6, an accuracy rate of 86%, Naïve Bayes showcases its ability to effectively classify passengers into satisfaction segments, contributing valuable insights for strategic decision-making in the airline industry. Regarding precision, which measures the accuracy of positive predictions, Naïve Bayes achieves a commendable score of 86%. This implies that the classifier adeptly identifies a significant proportion of true positive instances among all predicted positive cases, underscoring its precision in accurately categorizing satisfied passengers.

The recall metric, indicating the model's ability to capture true positive instances among all actual positive cases, is also reported at 86% for Naive Bayes. This highlights the classifier's effectiveness in recognizing satisfied passengers within the dataset, showcasing its capacity to capture positive instances without overlooking significant segments.

The F1-score, a holistic metric balancing precision and recall, is 86%. This composite measure emphasizes the equilibrium achieved by Naive Bayes in making accurate positive predictions while capturing a substantial proportion of satisfied passengers. Naïve Bayes, with its probabilistic approach, contributes meaningfully to customer segmentation efforts in the aviation sector. Further, the confusion matrix below explains the results more about how it falls on the testing dataset

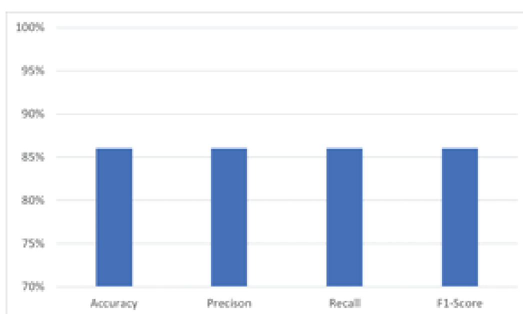

Figure 6: Performance Results with Naïve Bayes

In conclusion, Naïve Bayes performance across accuracy, precision, recall, and F1-score metrics positions it as a reliable classifier for uncovering patterns in the "Airline Passenger Satisfaction" dataset. While not surpassing the accuracy levels of specific other classifiers, Naïve Bayes provides a valuable probabilistic perspective for customer segmentation, offering nuanced insights to enhance satisfaction and retention strategies within the dynamic aviation landscape.

4.6. Conclusions

In conclusion, the findings of this study emphasize the significance of strategic customer segmentation in the airline industry. The classifiers exhibited commendable accuracy and efficiency, showcasing their applicability in identifying distinct passenger groups with specific preferences and behaviors. The practical implications highlight the potential for airlines to employ these classifiers in tailoring services, marketing strategies, and loyalty programs. The results affirm the value of integrating machine learning into customer relationship management, improving customer satisfaction and strengthening relationships. For future research, it is recommended to delve deeper into refining and expanding the dataset, incorporating additional features and diverse sources of information. Further investigation into ensemble methods and hybrid models could provide insights into optimizing classification performance. Exploring the integration of real-time data and continuous learning models can enhance the adaptability of the classifiers to evolving passenger preferences. Additionally, extending the study to encompass a broader range of airlines and demographic groups would contribute to a more comprehensive understanding of customer segmentation in the aviation sector.

REFERENCES

[1] I. Lewaaelhamd, "Customer segmentation using machine learning model: an application of RFM analysis", J*ournal of Data Science and Intelligent Systems,* vol. 2, no. 1, pp.29-36, 2024.

[2] V. R. Thalkar, "Customer Segmentation Using Machine Learning" *in International Journal of Scientific Research in Computer Science, Engineering and Information Technology,* ISSN DOI: https://doi.org/10.32628/CSEIT217654, 2023.

[3] S. Vikram et al., "Customer segmentation in tourism industry using machine learning models," I*nternational Research Journal on Advanced Science Hub,* vol. 5, no. 5, pp. 43-49, DOI: 10.47392/irjash.2023.S006, 2023.

[4] K. Tabianan et al., "K-means clustering approach for intelligent customer segmentation using customer purchase behavior data," *Sustainability,* vol. 14, no. 12, pp. 7243 DOI: 10.3390/su14127243, , 2022.

[5] Z. H. Sun et al., "GPHC: A heuristic clustering method to customer segmentation," *in Applied Soft Computing,* 111, pp. 107677, 2021.

[6] A. Griva et al., "A two-stage business analytics approach to perform behavioral and geographic customer segmentation using e-commerce delivery data," J*ournal of Decision Systems,* vol. 33, no. 1, pp.1-29, DOI: 10.1080/12460125.2022.2151071, 2024.

[7] J. M. Spoor, "Improving customer segmentation via classification of key accounts as outliers," *Journal of Marketing Analytics,* vol. 11, no. 4, pp. 747-760, DOI: 10.1057/s41270-022-00185-4, 2023.

[8] F. Meng et al., "Multiple dynamic pricing for demand response with adaptive clustering-based customer segmentation in smart grids," *In Applied Energy,* 333, pp. 120626, DOI: 10.1016/j.apenergy.2022.120626, 2023.

[9] P. Monil et al., "Customer Segmentation using Machine Learning," *In IJRASET,* ISSN: 2321-9653, IC Value: 45.98, SJ Impact Factor: 7.429, Vol. 8, no. 4, pp.2104-2108, Available at www.ijraset.com, June 2020.

[10] P. Kumar and A. K. Mokha, "Relationship between E-CRM, Customer Experience, Customer Satisfaction and Customer Loyalty in Banking Industry: A Review of Literature," *in RESEARCH REVIEW International Journal of Multidisciplinary,* vol. 6, no. 2, pp. 127-137, DOI: 10.31305/rrijm.2021.v06.i02.022, 2021.

[11] M. A. Gomes and T. Meisen, "A review on customer segmentation methods for personalized customer targeting in e-commerce use cases," *In Information Systems and e-Business Management,* DOI: 10.1007/s10257-023-00640-4, 2023.

[12] M. Loukili et al., "Sentiment Analysis of Product Reviews for ECommerce Recommendation based on Machine Learning," *In Int. J. Advance Soft Compu. Appl,* vol. 15, no. 1, DOI: 10.15849/IJASCA.230320.01, March 2023.

[13] T. L. Maheswari et al., "Customer Segmentation Based on Sentimental Analysis," *In 2022 International Conference on Advanced Computing Technologies & Applications (ICACTA),* pp. 1-7, DOI: 10.1109/ICACTA54488.2022.9753207, 2022.

[14] I. Pawełoszek, "Customer segmentation based on activity monitoring applications for the recommendation system," *In 25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems,* 192, pp. 4751–4761, DOI: 10.1016/j.procs.2021.09.253, 2021.

[15] M. Mashaabi, A. Alotaibi, H. Qudaih, R. Alnashwan, and H. Al-Khalifa, "Natural Language Processing in Customer Service: A Systematic Review," *in Computation and Language (cs.CL); Artificial Intelligence (cs.AI) , preprint arXiv:2212.09523*, 2023.

[16] E. Y. L. Nandapala and K.P.N Jayasena "The practical approach in Customers segmentation by using the K-Means Algorithm," I*n 15th (IEEE) International Conference on Industrial and Information Systems (ICIIS),* pp. 344-349, DOI: 10.1109/ICIIS51140.2020.934263, 2020.

[17] A. Alsayat, "Customer decision-making analysis based on big social data using machine learning: a case study of hotels in Mecca," *In Neural Computing and Applications,* vol. 35, no. 6, pp. 4701–4722, doi: 10.1007/s00521-022-07992-x, 2023.

[18] K. Tabianan et al., "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *in Data. Sustainability,* vol. 14, no. 12, pp. 7243, DOI: org/10.3390/su14127243, 2022,.

[19] P. S. Krishna and Dr. G. S. Sujatha, "Customer Segmentation Using Machine Learning, " *in © 2023 IJCRT,* vol. 11, no. 9, ISSN: 2320-2882, September 2023.

[20] J. K. Vieri et al., "Exclusive Clustering Technique for Customer Segmentation in National Telecommunications Companies," *In International Journal of Information Technology and Computer Science Applications (IJITCSA),* ISSN 2964-3139, vol. 1, no. 1, pp. 51 - 57, 2021.

[21] C. Dullaghan and E. Rozaki, "INTEGRATION OF MACHINE LEARNING TECHNIQUES TO EVALUATE DYNAMIC CUSTOMER SEG- MENTATION ANALYSIS FOR MOBILE CUSTOMERS," *In International Journal of Data Mining & Knowledge Management Process (IJDKP),* vol. 7, no.1, January 2017.

[22] A. Alghamdi, "A Hybrid Method for Customer Segmentation in Saudi Arabia Restaurants Using Clustering, Neural Networks, and Optimization Learning Techniques," *In Arabian Journal for Science and Engineering,* vol. 48, no. 2, pp. 2021–2039, DOI: 10.1007/s13369-022 07091.

[23] A. Ullah, M. I. Mohmand, H. Hussain, S. Johar, I. Khan, and S.Ahmad, "Customer Analysis Using Machine Learning-Based Classification Algorithms for Effective Segmentation Using Recency, Frequency, Monetary, and Time," *Sensors,* vol. 23, no. 6, pp. 3180, DOI: 10.3390/s23063180, 2023.

[24] P. Malviya et al., "Customer Segmentation and Business Sales Forecasting using Machine Learning for Business Development," *In International Journal on Recent and Innovation Trends in Computing and Communication,* DOI: 10.17762/ijritcc.v11i11s.8170, 2023.

[25] S. Vikram et al., "Customer Segmentation in Tourism Industry using Machine Learning Models," *In International Conference on intelligent COMPUting TEchnologies and Research (i-COMPUTER),* vol. 5, no. 5, pp.43-49, DOI: 10.47392/irjash.2023.S006, 2023.

[26] J. Joung and H. Kim, "Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews," *In International Journal of Information Management,* 70, pp.102641, DOI: 10.1016/j.ijinfomgt.2023.102641, 2023.