Analyzing factors influencing airline passenger satisfaction

Louis Fuego[1], James Andrei C. Nadela[3], Jomari Q. Celis[3]
College of Computing and Information Sciences, Mapua Malayan Colleges Mindanao, Davao City, Philippines[1]
College of Computing and Information Sciences, Mapua Malayan Colleges Mindanao, Davao City, Philippines [2]
College of Computing and Information Sciences, Mapua Malayan Colleges Mindanao, Davao City, Philippines[3]
Lcfuego@mcm.edu.ph[1]
janadela@mcm.edu.ph[2]
Jcelis@mcm.edu.ph[3]

Air travel is one of the most widely used forms of transportation, with approximately 4.5 billion people flying in 2024. To remain competitive, airlines need to understand passenger needs and improve satisfaction, a key factor in maximizing profits, enhancing service quality, and fostering customer loyalty. This study examines the factors influencing airline passenger satisfaction across the entire travel experience, from booking to arrival. The researchers tested four machine learning models—Naive Bayes, Decision Tree, Random Forest, and Gradient Boosting—to identify the best predictor of passenger satisfaction. Among these models, the Random Forest classifier emerged as the most effective, demonstrating a high balance of sensitivity, specificity, and accuracy while avoiding overfitting. These findings provide actionable insights for airlines aiming to enhance customer satisfaction and the overall passenger experience through targeted service improvements.

I. Introduction

1.1 Overview

One of the most popular forms of transportation is Air Travel. Approximately 4.5 billion people travel by air in 2024 [1]. Because of this demand, it is important for airline companies to know their passengers well. Airline passenger satisfaction is a key factor in determining whether an airliner succeeds in maximizing profits and offering a better service overall. Customer feedback is critical for business performance [2]. This study focuses on the satisfaction of a customer during the whole process. From Booking to arriving. Satisfaction is an important metric because it can influence a customer to be loyal and stay in the airline more [3].

While some studies have focused on specific methodologies, such as sentiment analysis utilizing natural language processing to discern key indicators of satisfaction [4], others adopt a broader perspective [5]. The latter approach emphasizes the influence of overall customer satisfaction on loyalty and the propensity of passengers to recommend an airline to others. This study will adopt this broader framework to investigate the relationship between customer satisfaction, brand loyalty, and the willingness of consumers to promote the airline to prospective travelers.

The study will Involve a dataset from Kaggle [6] "Airline Passenger Satisfaction" which has a total of 23 attributes and 25,975 rows.

1.2 Objective of the Study

Through the application of data science methodologies and machine learning algorithms, this study aims to make significant contributions to the existing literature in the following ways:

1. The research will conduct an in-depth analysis utilizing a dataset sourced from Kaggle, employing various machine learning models for comprehensive evaluation.

2. A comparative assessment will be performed to evaluate the efficacy of the most effective machine learning models and data science techniques applied to the specified dataset, emphasizing performance metrics and interpretability.

3. The study endeavors to predict passenger satisfaction by implementing advanced machine learning models, thereby providing insights into factors influencing user experience within the dataset context.

1.3 Concepts

Supervised Learning

Supervised learning encompasses two principal tasks: classification, which involves predicting categorical outcomes (such as in spam detection), and regression, which entails forecasting continuous numerical values (such as housing prices). During the training phase, the model employs a designated training set to acquire knowledge, while a separate testing set serves to evaluate the model's performance. Various metrics, including accuracy, precision, and mean squared error, are utilized to assess the model's predictive efficacy[10].

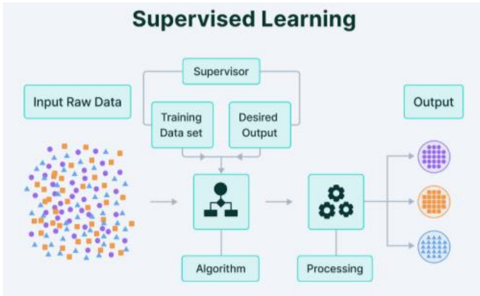

Figure.1.3.1

In the airline business, supervised learning can be applied to predict customer satisfaction by training a model on historical data that includes both customer feedback (such as satisfaction ratings) and a range of factors impacting their experience [11]. These factors, known as features, may include variables like flight delays, cabin crew service quality, seating comfort, food quality, check-in efficiency, and baggage handling. By using labeled data where each record links these features with a customer satisfaction score, a model can learn the relationships and predict satisfaction levels for future passengers.

The model could rate satisfaction using the Richter scale, with "1" being the lowest to "5" being the highest satisfaction level. The ratings will then be classified in different categories: "Neutral" "Satisfied" and "Dissatisfied". Insights gained from this model can help airlines identify which factors most significantly impact satisfaction, allowing them to make data-driven improvements. For instance, if the model shows that flight delays have a strong negative correlation with satisfaction, airlines may prioritize operational changes to improve punctuality. This application of supervised learning enables airlines to enhance the overall customer experience, increase loyalty, and potentially improve ratings and reviews.

Related Literature

A study analyzing 2,278 online passenger ratings worldwide used a multinomial logit model to assess the impact of service failures on the likelihood of promoting an airport. Findings indicate that failures related to airport staff and queueing times significantly reduce the probability of a positive recommendation, while issues with shopping and Wi-Fi are less impactful. Overall, the study suggests that service failures can greatly affect the airport experience, but passenger and airport characteristics do not significantly influence promoter likelihood [8]. A highly precise classification model with an AUC of 0.993, recall of 91.2%, and precision of 99.1% was developed to predict and enhance passenger satisfaction in the airline industry. The study recommends that airlines focus on improving in-flight Wi-Fi access and simplifying online booking to boost customer satisfaction and encourage loyalty, especially among first-time flyers [8]. Using logistic regression on the selected subset further clarified important variables, with findings indicating that online boarding and in-flight Wi-Fi significantly impact satisfaction across passenger and class types. Limitations include insufficient survey indicators and default model parameters, suggesting future enhancements should consider additional ground service metrics and optimized model tuning [9].

II. METHODOLOGY

In this section of the paper, the researchers will discuss the steps that are taken for this research. The researchers decided to take a page from the Data Science Process and followed the steps taken.

2.1 Data Collection and Acquisition

The first step taken was to acquire data. The dataset was taken from Kaggle, it consist of 25 columns which contains information about the customer's gender, age, travel type and various satisfaction levels of the airline's services.

Table Head	**Table Column Head**		
	Attributes	***Values***	***Data Type***
1	Gender	Male, Female	Nominal
2	Customer Type	Loyal, Disloyal	Nominal
3	Age	Ages Ranging from 7 – 80	Ratio
4	Travel Type	Business, Personal	Nominal
5	Class	Business, Eco, Other	Nominal
6	Flight Distance	Minutes Ranging from 31 to 4000	Ratio
7	Inflight Wifi Serive	Satisfaction Levels from 0 - 5	Ordinal
8	Departure/Arrial Time Convenience	Satisfaction Levels from 0 - 5	Ordinal
9	Ease of Online Booking	Satisfaction Levels from 0 - 5	Ordinal
10	Food and Drink	Satisfaction Levels from 0 - 5	Ordinal
11	Online Boarding	Satisfaction Levels from 0 - 5	Ordinal
12	Seat Comfort	Satisfaction Levels from 0 - 5	Ordinal
13	Inflight Entertainment	Satisfaction Levels from 0 - 5	Ordinal
14	On-board Service	Satisfaction Levels from 0 - 5	Ordinal
15	Leg Room Service	Satisfaction Levels from 0 - 5	Ordinal

Table Head	Table Column Head		
		Attributes / ***Values***	***Data Type***
16	Baggage Handling	Satisfaction Levels from 0 - 5	Ordinal
17	Check-in Service	Satisfaction Levels from 0 - 5	Ordinal
18	Inflight Service	Satisfaction Levels from 0 - 5	Ordinal
19	Cleanliness	Satisfaction Levels from 0 - 5	Ordinal
20	Departure Delay in Minutes	Satisfaction Levels from 0 - 5	Ordinal
21	Arrival Delay in Minutes	Satisfaction Levels from 0 - 5	Ordinal
22	Satisfaction	Neutral/Dissatisfied, Satisfied	Nominal

Table 2.1.1 List of Attributes after preprocessing

2.2 Data Cleaning and Pre-processing

Next, the researchers started with setting the roles for the variables in the dataset, where the "Satisfaction" column would be the label as this variable is what our model will be predicting. The researchers then looked at columns or variables that are of little to no significance when training the model. They then, looked for defects in the dataset, such as missing values, outliers, etc., that make the dataset unfit for processing and tried to fix those issues to improve the quality of the dataset.

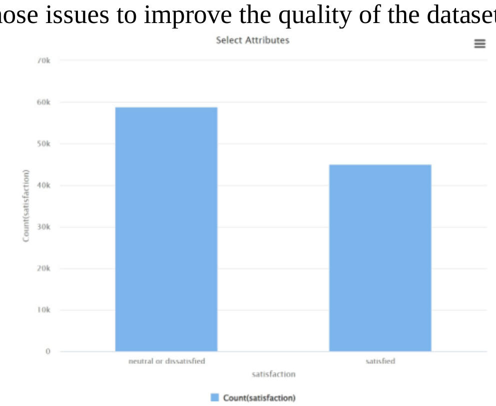

Figure 2.2.1 Sample Set after preprocessing

The researchers first checked for any imbalances in the dataset that they chose. The analysis shown in Fig 2.2.1 revealed a fairly balanced split, with almost 59k of passengers classified as neutral or dissatisfied and around 45k classified as satisfied. Given this near-even distribution between the two groups, no further adjustments, such as resampling or oversampling, were necessary to address class imbalance, allowing us to proceed without any problems.

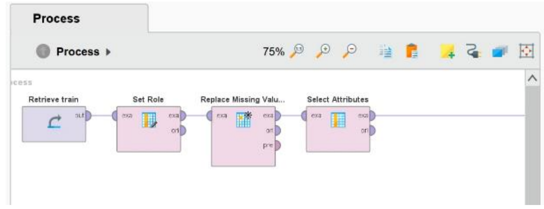

Figure 2.2.2 Preprocessing Subprocess

After which, **Label Encoding** was done which involves the process on setting roles for each variables in the dataset. The "Satisfaction" column was assigned as the label for this dataset. Next, any missing values were addressed by replacing them with the average value of the respective column to maintain data consistency. Following this **feature selection** was performed by excluding columns that did not contribute meaningful information to the model. Specifically, the "ID" column and the "Gate Location" column as they were irrelevant to the prediction task.

S`1

2.3 Data Transformation

Data Transformation is a step needed to take as it is important to improve the compatibility with various algorithms and enhance the model's performance.

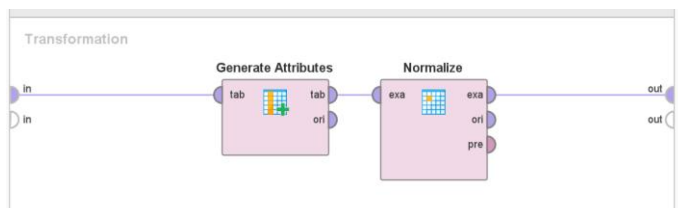

Figure 2.3.1 Data Transformation Subprocess

In this step, the researchers began the data transformation process by first grouping attributes through conditional statements. The column age was categorized into different age groups: 0-17, 18-24, 25-34, 35-44, 45-54, 55-64 and 65+.

Then the researchers applied the normalization of our data, specifically the "Flight Distance" attribute. It was then normalized through z-transformation. The reason for this normalization was to aim for a more consistent data and to avoid biases with the flight distance. [9]

III. DATA ANALYSIS

In order to perform analysis and create a prediction model with the data set, classification needs to be performed. The researchers will be using different classification algorithms including Decision Trees, Random Forest, Naïve Bayes, and Gradient boosted trees to see which algorithm has the best overall performance in classifying customer satisfaction.

3.1 Chosen Algorithms:

Decision trees: This algorithm handles categorical data well making it well suited for this classification model [12]. The data that this algorithm produce is also easily interpretable making it easier to extract meaningful insights from the results.

Random Forest: This algorithm utilizes multiple decision trees to produce more accurate results [13]. This algorithm is also ideal for classification as it reduces overfitting which is where the algorithm can't produce accurate predictions from new data. However, it is more computationally extensive than the decision tree leading to longer training times.

Naïve Bayes: This is the simplest among the listed algorithms making it computationally efficient, where training times are faster compared to decision tress. This algorithm also works well with categorical data which suits the given dataset [14].

Gradient Boost Trees: This algorithm can sometimes be more accurate compared to a random forest due to its training method [15]. It builds a sequence of models where each learns from the mistake of the previous model leading to higher accuracy.

3.2 Procedure:

After preprocessing the data, the classification model was then created. In rapid miner there are two ways to do that, one is by using the auto model feature which automatically tests and trains the model. This auto model provides analytics and the performance of the models selected. The other way is to utilize the drag and drop feature in the design panel. In here, the data was split into training data (70%) and testing data (30%). The training data will then be used to train the model and the testing data will be used to test the model's performance by introducing new data the model has not seen. For the rapid miner the multiply operator was used to allow training for different models at one design panel. For the models, the base parameters were used to train the classification models.

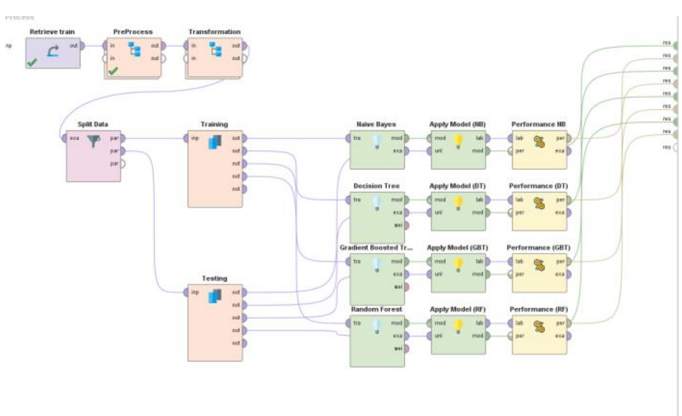

Figure 3.2.1 Processing data

3.2.1 Model Parameters:

The algorithms were set to ensure that the models' reliability and accuracy

- Naive Bayes: Used the default parameters in rapid miner
- Decision tree: Was set to have a maximal depth of 20.
- Gradient Boosetd Trees: Was set to have 50 trees and a maximal depth of 20.
- Random Forest: Was set to have 100 number of trees and a maximal depth of 20.

3.3 Exploratory Data Analysis

By examining the feature weight from the random forest results, insights can be gained into identifying factors that greatly influence customer satisfaction.

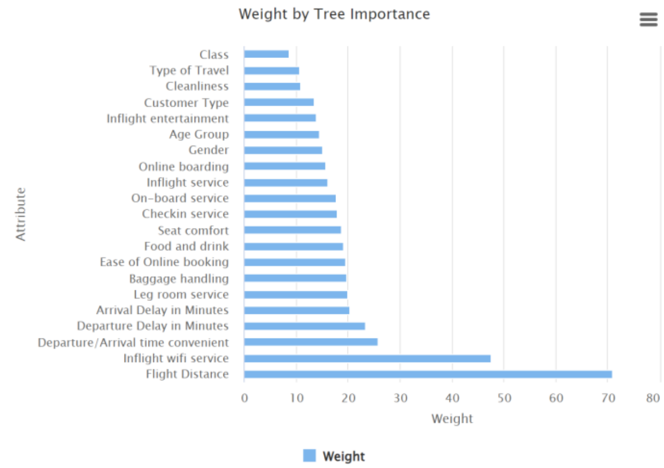

Figure 3.4.1 weights by features

From the figure above, the Flight distance had the most significant impact on predicting customer satisfaction, followed by inflight services and departure/arrival time convivence. Conversely, class, gender, and age groups had less of an impact. This graph suggest that factors related to flight had more of an impact compared to the customer demographic.

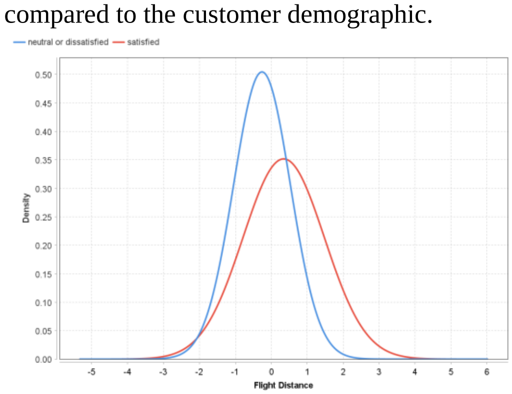

Figure 3.4.1 flight distance bell curve (normalized)

From the graph reveals that shorter flights tend to have more dissatisfied/neutral customers, while longer flights tend to have a higher proportion of satisfied customers. This could be attributed to improved inflight amenities like entertainment, seat comfort and meals.

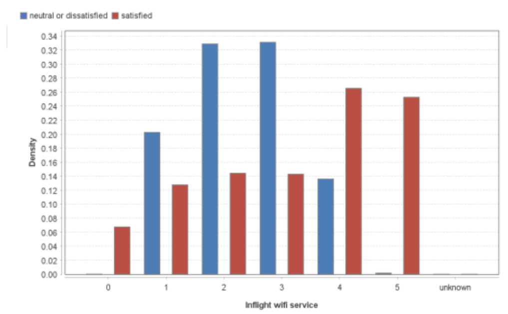

Figure 3.4.2 inflight Wi-Fi service

The graph at figure 3.4.2 reveals that a lower rating had more unsatisfied customers. On the other hand, a higher rating on the Wi-Fi service led to more satisfied customers with customers who rated the service a 5 being almost all satisfied.

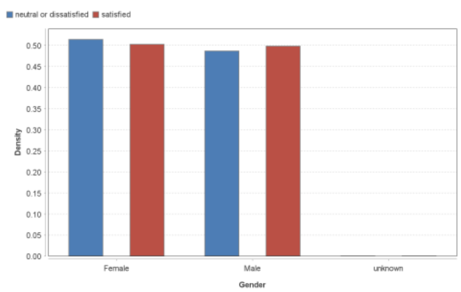

Figure 3.4.3 gender demographic

The graph above shows how there is little difference between the satisfaction levels of males and females.

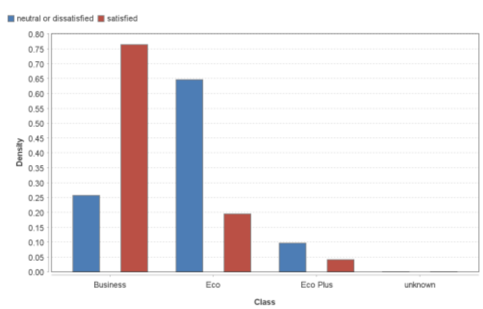

Figure 3.4.4 class distribution

The class distribution graph shows that customers on business class tend to be more satisfied while customers on eco class tend to be more neutral/dissatisfied.

3.3 Model Evaluation metrics

To assess the performance of each model several evaluation metrics were used including:

1. Accuracy: a metric on how well a model can predict the right outcomes.
2. Sensitivity: correctness of a model in identifying true positives or how well it can correctly identify satisfied customers.
3. Specificity: correctness of a model in identifying true negatives or how well it classifies neutral or dissatisfied customers.
4. False Positive Rate: Portion of negative cases that are incorrectly identified as positive

IV. RESULTS AND DISCUSSION

4.1 Model Performance and Analysis

	Sensitivity	Specificity	Accuracy	TP Rate	FP Rate
Naive Bayes	82.58%	86.87%	85.01%	82.58%	13.13%
Decision Tree	93.29%	96.50%	95.10%	93.29%	3.50%
Random Forest	93.88%	97.26%	95.79%	93.88%	2.74%
Gradient boosted tree	94.56%	95.69%	95.35%	94.56%	3.94%

Table 4.1.1 Performance

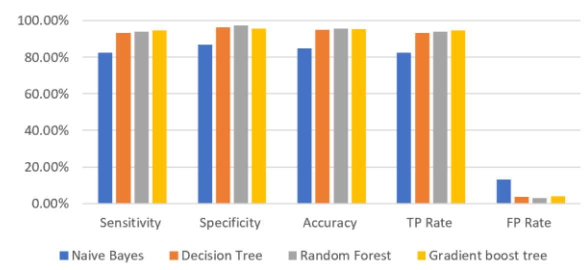

Figure 4.1.2 model performance graph

The evaluation of each model shows some key insights:

- Naive Bayes: Compared to the other models, this model has the lowest scores on all metrics due to it being the simplest algorithm. The FPR is also significantly higher indicating that it detects more incorrect cases.
- Decision tree: Even though this algorithm is less complex than the gradient boosted tree and Random Forest it was still able to have similar scores. This is due to the dataset being simple where features had straightforward relationships.
- Random Forest: This model had the best accuracy indicating its ability to generalize well to unseen data and make reliable predictions.
- The Gradient Boosted Tree model had slightly lower scores than the random forest except for the sensitivity, where it scored the highest indicating that it is able to accurately classify positive cases.

The assessment of each model provides important findings for the model's strengths and areas to improve on.

V. Conclusion and Recommendation

Based on our analysis, the researchers recommend the Random Forest model as the best choice for this study. This classifier demonstrates a well-balanced performance across sensitivity, specificity, and accuracy among all models tested. Specifically, it achieves a sensitivity of 93.88%, indicating a high true positive rate, and a specificity of 97.26%, which reflects strong accuracy in identifying true negatives. With an overall accuracy of 95.79% and a low false positive rate of 0.0394, the Random Forest model consistently provides reliable predictions across both positive and negative classes. These results surpass those of Naive Bayes, Decision Tree, and Gradient Boosting, making Random Forest the most effective model for minimizing misclassifications and enhancing predictive reliability in our study.

The Gradient Boosted Tree Model had comparable scores to the other models, outperforming Decision Trees on most metrics while excelling in sensitivity. It showed high accuracy in classifying positive cases. To further enhance the performance of all models evaluated, researchers should concentrate on optimizing preprocessing techniques such as feature selection and scaling, as well as implementing strategies to mitigate overfitting where relevant. Limiting tree depth, the usage of early stopping, and the application of regularization are techniques that can improve the reliability and generalization of models like Gradient Boosting and Decision Tree, while hyperparameter tuning and cross-validation can help all models reach their full potential in accurately classifying new data.

In conclusion, our analysis identified the top four factors influencing airline passenger satisfaction: Flight Distance, Inflight Wi-Fi Service, Departure Delay in Minutes and Departure/Arrival time convenience. Among these, Inflight Wi-Fi Service is the 2nd most influential factor behind flight distance. These findings, combined with the high predictive power of the Random Forest model, highlight the value of focusing on these specific service aspects to enhance passenger satisfaction. Improving these factors, particularly Inflight Wi-Fi, could yield significant benefits for passenger experience, as indicated by our model's reliable insights into key satisfaction drivers.

References

[1] Statista. (2024, October 11). *Global air traffic - scheduled passengers 2004-2024.* https://www.statista.com/statistics/564717/airline-industry-passenger-traffic-globally/

[2] Halpern, N., & Mwesiumo, D. (2021). Airport service quality and passenger satisfaction: The impact of service failure on the likelihood of promoting an airport online. *Research in Transportation Business & Management*, *41*, 100667. https://doi.org/10.1016/j.rtbm.2021.100667

[3] Ban, H., & Kim, H. (2019). Understanding Customer Experience and Satisfaction through Airline Passengers' Online Review. *Sustainability*, *11*(15), 4066. https://doi.org/10.3390/su11154066

[4] Eshaghi, M. S., Afshardoost, M., Lohmann, G., & Moyle, B. D. (2024). Drivers and Outcomes of Airline Passenger Satisfaction: A Meta-Analysis. *Journal of the Air Transport Research Society*, *3*, 100034. https://doi.org/10.1016/j.jatrs.2024.100034

[5] J. Roller, "Why is Data Normalization Important?," *IEEE Computer Society*, Mar. 04, 2024. https://www.computer.org/publications/tech-news/trends/importance-of-data-normalization/

[6] Hulliyah, K. (2021). Predicting Airline Passenger Satisfaction with Classification Algorithms. IJIIS International Journal of Informatics and Information Systems, 4(1), 82–94. https://doi.org/10.47738/ijiis.v4i1.80

[7] Jiang, X., Zhang, Y., Li, Y., & Zhang, B. (2022). Forecast and analysis of aircraft passenger satisfaction based on RF-RFE-LR model. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-14566-3

[8] *What is Supervised Learning? | Google Cloud*. (n.d.). Google Cloud. https://cloud.google.com/discover/what-is-supervised-learning

[9] Ibrahim, A. O., Yi, C. C., Elsafi, A., & Ghaleb, F. A. (2024). Revolutionizing Airline Customer Satisfaction Analysis with Machine Learning Techniques. In *Lecture notes on data engineering and communications technologies* (pp. 141–152). https://doi.org/10.1007/978-3-031-59707-7_13

[10] GeeksforGeeks. (2024, May 17). Decision tree. GeeksforGeeks. https://www.geeksforgeeks.org/decision-tree/

[11] GeeksforGeeks. (2024, July 12). Random Forest algorithm in machine learning. GeeksforGeeks. https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/

[12] Prabhakaran, S. (2022, April 20). *How Naive Bayes Algorithm Works? (with example and full code)*. Machine Learning Plus. https://www.machinelearningplus.com/predictive-modeling/how-naive-bayes-algorithm-works-with-example-and-full-code/

[13] Tuychiev, B. (2023, December 27). A guide to the gradient boosting algorithm. https://www.datacamp.com/tutorial/guide-to-the-gradient-boosting-algorithm

[14] An, M., & Noh, Y. (2009). Airline customer satisfaction and loyalty: impact of in-flight service quality. Service Business, 3(3), 293–307. https://doi.org/10.1007/s11628-009-0068-4

[15] Rajendran, S., & Srinivas, S. (2021). Air taxi service for urban mobility: A critical review of recent developments, future challenges, and . . . ResearchGate. https://doi.org/10.48550/arXiv.2103.01768