Paper Title (*Utilizing Machine Learning To Forecast Passenger Satisfaction in Airline Industry*)

Authors Name/s per 1st Affiliation *(Author)*
line 1 (of *Affiliation*): Business Intelligence Program
line 2: Hogskolan Dalarna University
line 3: Borlange, Sweden
line 4: e-mail: v24pujkh@du.se

Authors Name/s per 2nd Affiliation *(Prayag Narasingh Rana)*
line 1 (of *Affiliation*): Business Intelligence Program
line 2: Hogskolan Dalarna University
line 3: Borlange Sweden
line 4: e-mail: v24prara@du.se

***Abstract*— The proposed endeavor aims to examine the application of machine learning algorithm for predicting passenger satisfaction within the airline business, addressing the up-surge demand for data-driven service optimization. A publicly accessible dataset comprising travel information, consumer demographics, and service evaluations was employed. Random forest classifier, Decision tree, and logistic regression is employed in the proposed endeavor. The Random forest model is able to produce high accuracy of 92% with high precision, recall and F1-Score vales. Utilizing the model, feature determinants of customer satisfaction was identified such as Online boarding, seat comfort, and in-flight entertainment. The outcomes show how efficient the model is and offer insightful information for enhancing the airline industry's customer experience.**

Keywords- component; formatting; style; styling; Airline customer satisfaction, Machine Learning, Random forest Classifier, Preductive modeling, service quality, Frature Importance, Decision tree, Logistic Regression, Satisfcation.

I. INTRODUCTION *(HEADING 1)*

Competition among the airlines company was been intensified due to rapid growth of the airline industry, forcing these companies to have up-surge emphasis on the customer satisfaction by(Jindal Krish et al., 2024). According to (Murugesan R. et al., 2024) it vital to comprehend factors affecting customer satisfaction for enhancing the service quality, remain competitive among the airline business and ensure the customer retentions. However, identifying the factors that influence the customer and measuring satisfaction effectively has been challenging to the business, given the subjective human nature experiences and the infinite number of service and the demographic data gathered by the airline(Luo, 2023).

This endeavor aims to address the problem of predicting the airline customer satisfaction utilizing machine learning. Particularly, using Random Forest classification model this endeavor aims to develop and analyze whether the customers are satisfied or unsatisfied depending on the different service factors and features of demographics. Airline companies thus can target the areas for improvement of their services, optimize the allocation of recourses and upsurge the overall customer experience with the help of the actionable insight from this problem (Baswardono et al., 2019).

This project is centered around the following key research questions:

Research question 1: How well can airline passenger's satisfaction can be predicted using the Random Forest model?

Research question 2: Which are the key features in airline that impact the passenger satisfaction?

This project mainly assesses predictive performance of Random Forest classifier by comparing the model with two other model and analyze the most influential features that determine the customer satisfaction. For the data processing and analysis of the model's performance, key metrics has been accessed such as accuracy, precision, recall, F1-score and feature importance. We utilized the data from publicly available dataset from Kaggle (Ramin, 2024).

This report adds to the existing research to enhance the customer experiences in the airline business by utilizing the machine learning technique in a practical context. Random Forest classification model facilitates high performing feature importance analysis. Resilience to overfitting and power to manage high dimensional data, ensuring its most flexible for complex, practical dataset for the airline customer satisfaction is the main motivation to utilize the Random Forest model.

II. LITERATURE REVIEW

1. Predicting customer satisfaction importance

Customer satisfaction is vital metric to directly impact the business performance such as brand image, market share and the customer loyalty. Due to increased competition and need for data driven decision making, airline sector is under pressure to upsurge the customer experience, particularly in the post-pandemic. According to(Li, 2025) entertainment within flight, services provided by the cabin and seat comfort are key aspect of travel that is closely connected to the customer satisfaction.

The dataset containing over 129,000 records was utilized for the proposed endeavor which includes the information regarding service and the demographic. The report also emphasized that the demographic characteristic such as distance and age have certain role while most impacting factors are service-related features like ticket booking application, environment quality in flight, online customer service and so on. This insight highly supports our endeavor objective to use service and demographic related data to make prediction for customer satisfaction levels utilizing techniques of machine learning.

2. Random forest for predictive modeling

Among several machine learning models, Random Forest classification model, is one of the most pertinent contribution of Li's research to our project. The research has utilized different model's such as Logistic Regression, Random Forest and Neural Network, among which Random Forest shows the highest accuracy with 95.87% and AUC-ROC 0.9937, ensuring its capacity to robustness in managing non-linear relationship, complexity in the data.

The Random Forest model is also provided with the ability to generate feature importance which rank the factors affecting the customer satisfaction like entertainment in flight, seat comfort, cabin service and so on, which enables the business to priorities the feature with highest ranking. This closely related to our project goal, which are to pinpoint and measure the main factors influencing satisfaction. To predict and generate the actionable insights for business to comprehend the feature importance, this model is most beneficial.

Furthermore, Random Forest classification model demonstrated effectiveness in handling large and multidimensional dataset and resilience's to overfitting which is especially helpful when working with the publicly assessable dataset such as Kaggle.

3. Implications of airline service strategy

From Li's research, we can extract clear guidance for enhancement of airline service strategies. More focused should be drawn towards inflight experiences such as enhancing the entertainment system, improvement of the seat ergonomics, guaranteeing consistent cabin services and so on. Even tradition attributes are still relevant like pricing and punctuality, they seem to have correlation with satisfaction in the present experience -based travel context.

These outcomes reinforce the rationale for encompassing on experience-based attributes in the modeling phase for our project. nevertheless, the performance metrics use such as accuracy, recall, precision, F1-score and feature importance justify the model effectiveness and enables to adjust practical services for the airline business.

III. METHOD DESCRIPTION

A. The Dataset (Heading 2)

For the proposed project, airline customer satisfaction dataset was utilized from publicly assessable dataset "Kaggle" (Ramin, 2024). The dataset consists of 129,880 records and 23 different attributes associated with the quality of service, demographic, flight characteristics and customer satisfaction. The dataset's column consists of **"satisfaction"** which provide the information regarding customer's satisfaction level, **"customer Type"** which categories the customers as **"Loyal"** or **"Disloyal".** The dataset also includes the demographic information such as column of customer ***"Age"*** which describe age of particular customer. Details of customers travel is described by the column **"Type of Travel"** which consists of what type of ticket does the customer purchases while travelling such as **"Business"** or **"personal"** and under the **"class"** column **ticket class such as "Economic or Eco", "Business" or "Eco Plus"** are mentioned. **"Flight Distance"** column provide the distance travelled in kilometers are recorded.

Additionally, the dataset consist rating across multiple dimensions which are rated from 1 to5 like "**Seat Comfort, Time convenience of Departure/ Arrival, quality of Food and Drink, Location of gate, Wi-Fi services within the flight, Online support, Onboard service, ease while making online bookings, leg room service, handling of baggage, check-in services, online boarding and cleanliness".** Furthermore, the dataset also covers the information on flight punctuality which are recorded under **"Departure Delay in Minutes"** and **"Delay of Arrival in Minutes".** Utilizing this dataset enables us to analyze detailed factors impacting the level of satisfaction of customer in the airline business.

Expect for the Arrival Delay column, the dataset in hand was clean from the beginning, while the Arrival Delay column consist of some NA values. Some of the attributes such as (**Departure Delays in Minutes, Flight Distance and Gate Location)** in the dataset was negatively correlated, so we dropped these columns to reduce noise in data and enhance the model focus on relevant predictors along with the column **Arrival Delay in Minutes which is displayed by the figure 1.**

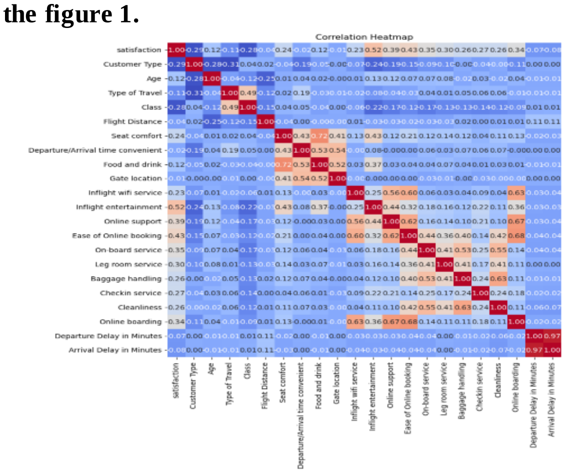

Figure 1 Correction matrix showing negative correlation

Before generating the correlation matrix, we change categorical column such as **"Customer Type", "Type of travel" "Class**" and **"Satisfaction"** to numerical using onehot encoding. It converts categorical variables of elements to number systems such as **“Customer Type”** have two different variables **“loyal”** to 1 and **“Disloyal”** to 0, in column **“Class”** it has 3 attributes, 0 for **“Economic”**, 1 for **“Eco Plus”** and 2 for **“Business”** and so on. Also, the correlation matrix generated after dropping the non-necessary columns is demonstrated by the figure 2.

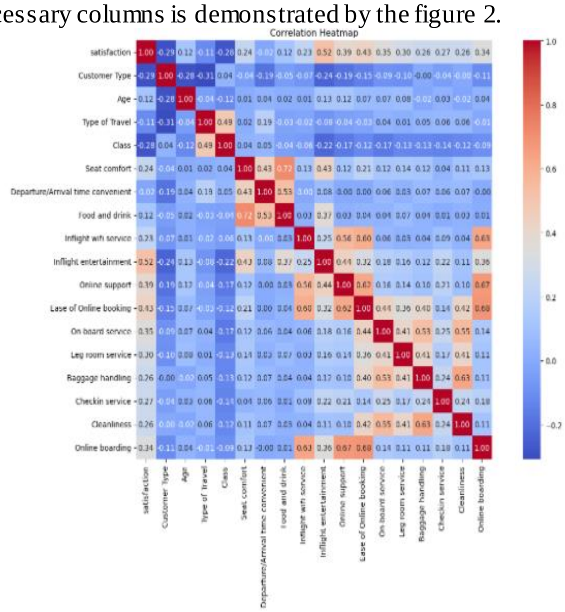

Figure 2 Correlation matrix after removing the negatively correlated column

B. Data Mining Method

In order to predict satisfaction of airline customer, we utilized Random forest Classifier, Decision tree and Logistic regression. We used three models to comprehend which of the model is best for predicting the airline customer satisfaction.

The target variable **“Satisfaction”** was separated from the predictors. We split the dataset into 80% training and 20% testing sets to provide an objective assessment of the data. This fixed-random split ensured reproducibility of results. The data was then divided as “X” and “Y” variables, where “X “being the training dataset, “Y” is the target data.

Model training

1. **Random Forest:** We instantiated a Random Forest Classifier with n_estimators = 100, a maximum depth of 10, a minimum of 10 samples to split an internal node and a minimum sample per leaf= 5. These fixed hyperparameters along with random_state=42 ensure the fitting in the data.
2. **Decision tree:** using the same training data, overfitting-control parameters was utilized in decision tree classifier with max_depth=5, min_sample_split=10, min_sample_leaf=5 and Random_state=42.
3. **Logistic regression:** Utilizing “Liblinear” solver for convergence on the same trained data, Logistic Regression was utilized. Defult regularization with randsom_state =42 was utilized for consistency.

One of the benefits of Random forest’s was the capability to measure feature importance which provide list important features that mainly influence the customer’s satisfaction which enables to discover key drives and thus generate actionable insights for enhancement of airline services. We, therefore access the importance features driving the customer’s satisfaction.

For more robust evaluation, we utilized a stratified K-fold cross-validation technique. The dataset was divided into 5 folds, where each of the folds includes a class distribution like the original dataset. For instance, if the dataset consists of 60% “satisfied” and 40% “not satisfied”, every fold reflects the same ratio. From the 5 folds, 4 folds are utilized as training while 1fold is for the testing. after the completion of all 5 iteration, the metrics of evaluation is averaged across all folds. A more accurate measure of the efficiency of generalization is provided by this method, which guarantees that each point of data gets utilized for validation as well as training. The model performance was accessed on the test set and across cross validation folds as Accuracy score, precision, recall, F1-score, ROC-AUS Score, Confusion Matrix.

Lastly, we utilize McNemar’s test to test whether the difference in performance among the model were statistically significant. As a baseline we created dummy classifier which always predicted the most frequent class. Predictions for comparison of each pair of models (Random forest vs. baseline, Random forest vs. logistic regression) were obtained on the test dataset, which were then compared to actual labels to generate binary arrays showing whether each prediction was correct or incorrect.

IV. RESULTS AND ANALYSIS

The models were assessed based on their capacity to predict satisfaction of airline customer, a binary classification problem in which the goal variable indicates whether a customer is “satisfied” or “not satisfied”. With high accuracy of 92.97% Random Forest classifier demonstrate strong capacity of predictive performance on test set which pin point the model’s effectiveness at classifying satisfied and dissatisfied customer depending on service and experience-based features.

In contrast, Random Forest surpasses both models on important measures, with the Decision Tree Classifier achieving a test accuracy of 87.63% and Logistic Regression achieving

we generated metric scores to compare the best fitting model. We generated precision, recall, f1-score and the accuracy of the models. We also compared the models in a K-fold cross validation and drew conclusions using the ROC Curve. Furthermore, Confusion Matrix was generated for all three models in order to evaluate the performance of classification models by showing the number of correct and incorrect predictions for each class.

The results are as follows:

Random Forest Classifier:

```
Random Forest Classifier:
Training Accuracy: 0.9298
Test Accuracy: 0.9297
              precision    recall  f1-score   support

           0       0.93      0.92      0.92     11675
           1       0.93      0.94      0.94     14301

    accuracy                           0.93     25976
   macro avg       0.93      0.93      0.93     25976
weighted avg       0.93      0.93      0.93     25976
```

Figure 3 Evaluation metric for Random forest classifier showing accuracy, precision, recall and f1 score

Random Forest Classifier showed a training accuracy of 0.92 equal to the testing accuracy of 0.92. The model showed a strong predictive performance with minimal overfitting. This suggests that the model generalizes well for data that is unseen. The model performed well in both classes. The precision score of both “Satisfied” and “Not Satisfied” was 0.93. The recall for “Not Satisfied (0)” was 0.92 and “Satisfied (1)” was 0.94. The balanced precision and recall scores indicate that the model is not biased towards the majority class. F1-score of 0.92 on Not Satisfied and 0.94 on Satisfied (both high) shows that both classes are well-controlled.

Random Forest <u>Confusion Matrix</u>

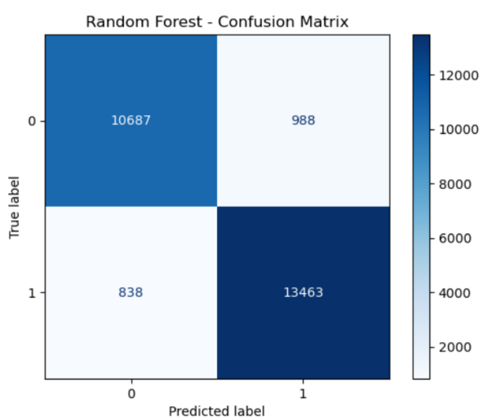

Figure 4 Confusion Matrix for Random Forest Classifier showing the count of True Negative, False, False Negative, True Positive

Table 1 Confusion Matrix table for Random Forest showing count of True Negative, False Positive, False Negative, True Positive

	Count	Description
True Negative	10,687	Correctly predicted 'Not Satisfied'
False Positive	988	Predicted 'Satisfied' when actually 'Not Satisfied'
False Negative	838	Predicted 'Not Satisfied' when 'Satisfied'
True Positive	13463	Correctly predicted 'Satisfied'

Random Forest Model shows a strong performance because it correctly classifies 10,687 cases as ‘Not Satisfied’ (True negatives) and 13,463 as ‘Satisfied’ (True Positives). Random Forest also classified 988 False Positives i.e. ‘Satisfied’ instead of ‘Not satisfied’ and 838 False Negatives i.e. ‘Not satisfied’ instead of ‘Satisfied’.

Decision Tree Classifier:

```
Decision Tree Classifier:
Training Accuracy: 0.8749
Test Accuracy: 0.8763
              precision    recall  f1-score   support

           0       0.85      0.89      0.87     11675
           1       0.90      0.87      0.89     14301

    accuracy                           0.88     25976
   macro avg       0.87      0.88      0.88     25976
weighted avg       0.88      0.88      0.88     25976
```

Figure 5 Evaluation metric for Decision Tree showing accuracy, precision, recall and f1 score

The decision tree shows a training accuracy of 0.87 and test accuracy of 0.87. ‘Not Satisfied (0)’ of 0.85 and ‘Satisfied (1)’ of 0.90 on the precision metric. 0.89 “Not Satisfied (0)” and 87% Satisfied (1) on recall metric. Both precision and recall are slightly lower than those of the Random Forest. This shows that the model is more likely to misclassify ‘Not Satisfied’.

Decision Tree <u>Confusion Matrix</u>

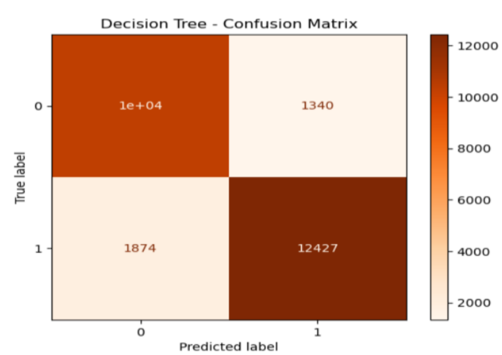

Figure 6 Confusion Matrix for Decision Tree Classifier showing the count of True Negative, False, False Negative, True Positive

Table 2 Confusion Matrix for Decision Tree table showing count of True Negative, False Positive, False Negative, True Positive

	Count	Description
True Negative	10335	Correctly predicted 'Not Satisfied'
False Positive	1340	Predicted 'Satisfied' when actually 'Not Satisfied'
False Negative	1874	Predicted 'Not Satisfied' when 'Satisfied'
True Positive	12427	Correctly predicted 'Satisfied'

Compared to Random Forest, Decision tree performed even less effectively with 352 more False Positives and about 1,035 more False Negatives than Random Forest. This shows that Decision tree has an overall less ability to capture both classes.

Logistic Regression

```
Logistic Regression Accuracy: 0.8254

Classification Report:
              precision    recall  f1-score   support

           0       0.80      0.81      0.81     11675
           1       0.84      0.84      0.84     14301

    accuracy                           0.83     25976
   macro avg       0.82      0.82      0.82     25976
weighted avg       0.83      0.83      0.83     25976
```

Figure 7 Evaluation metric for logistic regression showing accuracy, precision, recall and f1 score

This model shows a moderate performance on both classes, but it is slightly biased towards correctly predicting the 'Satisfied' class. The accuracy for Logistic Regression is 0.82. It has a precision score of 0.80 on "Not Satisfied (0)" and 0.84 on "Satisfied (1)". The recall score of "Not Satisfied" is 0.81 and "Satisfied" is 0.84. Although precision and recall are decent, they are noticeably lower than those of the previous complex tree-based models.

Logistic Regression Confusion Matrix

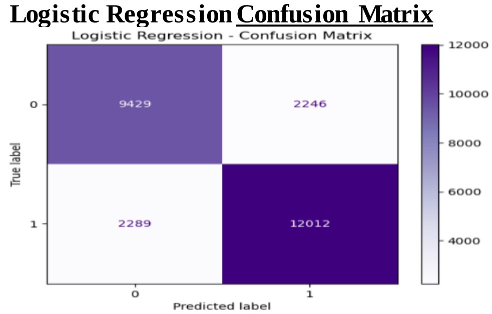

Figure 8 Confusion Matrix for Logistic Regression showing the count of True Negative, False, False Negative, True Positive

Table 3 Confusion Matrix table for Logistic Regression showing count of True Negative, False Positive, False Negative, True Positive

	Count	Description
True Negative	9429	Correctly predicted 'Not Satisfied'
False Positive	2246	Predicted 'Satisfied' when actually 'Not Satisfied'
False Negative	2289	Predicted 'Not Satisfied' when 'Satisfied'
True Positive	12012	Correctly predicted 'Satisfied'

Amongst the three, logistic regression showed the least reliability i.e. weakest performance. Both False Negatives and False Negatives were highest, and it had a poor performance especially identifying "Not Satisfied".

Feature importance

Using the benefits of Random Forest classifier, features which contributed the most to predict model was calculated called as feature importance.

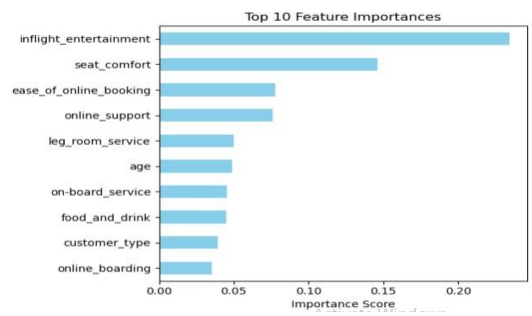

Figure 9 Horizontal bar graph showing feature importance of the dataset

The results obtained are in line with those of (Li, 2025) and (Sobowale, 2024) who both found that comfort and in-flight amenities were the main determinants of satisfaction. The little contribution of less important characteristics, like flight distance, gate location, and arrival delay, confirmed that they were eliminated during pre-processing.

Stratified K-Fold with Shuffling

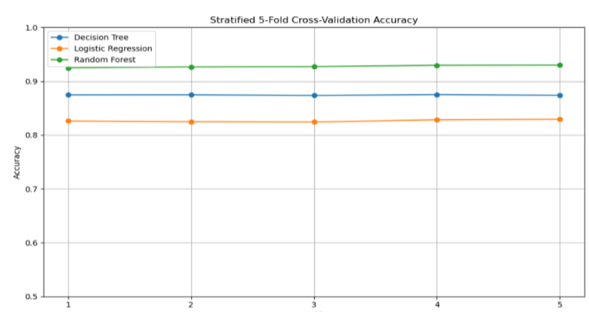

Figure 10 Stratified K-Fold Cross Validation showing accuracy of all three models across each fold

All three models show very high consistency and stability in all 5 folds.

Random Forest has the highest accuracy across all folds showing its reliability. Low SD of 0.0019 shows the minimal variance meaning no fold experienced unusual drops.

Decision Tree shows and even more stable performance than Random Forest but the mean accuracy is significantly lower (0.0006). This shows that Random Forest is more powerful over even though the Decision tree is more consistent.

The mean accuracy of Logistic Regression is 82.66% i.e. it clearly underperforms compared to both the tree models. SD of 0.002 indicates it has slightly more variation but still acceptable.

ROC Curve

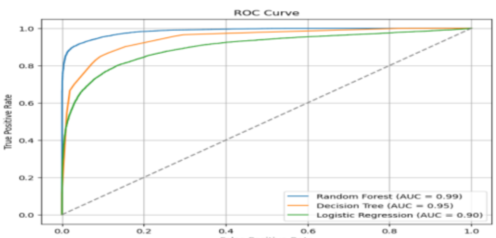

Figure 11 ROC Curve showing model stability and distinguishability with Area Under Curve for all three models

In the ROC curve we can see that Random Forest performs much better than Decision Tree and Logistic Regression. With an Area Under Curve (AUC) of 0.99, Random Forest shows almost perfect stability and an ability to distinguish between satisfied and not satisfied customers. Decision Tree also performs well with an AUC of 0.95. Logistic Regression manages to get an AUC of 0.90.

These results support the conclusion that Random Forest is the most effective model for this binary classification. From the results above we can conclude that Random Forest is better suited to predict the Satisfied and not Satisfied classes for this dataset.

Statistical comparison (McNemar's Test)

To analyze if the differences in classification accuracy among models are statistically important, we utilized McNemar's Test.

Table 4 McNemar's statistical testing showing comparison of three models

Model comparison	Correct by A only	Correct by B only	p-value
Random Forest vs baseline	10,687	838	0.0
Decision Tree vs Baseline	10,335	1,874	0.0
Logistic Regression vs Baseline	9429	2,289	0.0
Random Forest vs Decision Tree	1,7987	399	~0.0
Random Forest vs Logistic Regression	3,139	430	0.0

P-values were less than 0.05 in every comparison, showing statistically significant variations in predicting performance. The Random Forest Classifier was the most resilient and dependable model in this investigation, consistently outperforming the baseline and simpler models.

I. CONCLUSION

The main objectives of this project were to develop a predictive model which can classify the satisfaction of the airline customers utilizing machine earning technique. after a thorough analysis, we identified that Random Forest classifier was the most effective model for predicting airline customer satisfaction. With highest accuracy, strongest ROC_AUC performance and the lowest error rates, Random forest classifier outperform other models such as Decision Tree and Logistic Regression.

With the help of feature importance, we were able to identified features that are most influential in measuring the satisfaction of airline. Model's robustness and capability to generalize to unseen data was ensured thought different evaluation such as confusion matrices, classification metrics, ROC curves and stratified 5-fold Cross – Validation. Additionally, with the help of McNemar's Test we were also to perform statistical testing that verified that the performance enhancement of the Random forest model is statistically vital when compared to alternative models and the baseline.

To summarize, Random Forest model shows excellent prediction capability and usefulness to actual airline customer satisfaction.

REFERENCES

[1] Baswardono, W., Kurniadi, W., Mulyani, A., & Arifin, D. M. (2019). Comparative analysis of decision tree algorithms: Random forest and C4.5 for airlines customer satisfaction classification. *Journal of Physics: Conference Series*, *1402*(066055). https://www.academia.edu/112384390/Comparative_analysis_of_decision_tree_algorithms_Random_forest_and_C4_5_for_airlines_customer_satisfaction_classification

[2] Jindal Krish, Er. Sharma Disha, Gupta Akrit, Sharma Pranjal, & Mansha, G. (2024). Airline Customer Satisfaction Prediction. *Multidisciplinary Research* *6*(2). https://www.ijfmr.com/papers/2024/2/18640.pdf

[3] Z. (2025). Analyzing Key Factors Influencing Airline Customer Satisfaction Using Machine Learning Models. *Financial Technology and Business Analysis* *155*(1), 67-76. https://doi.org/10.54254/2754-1169/2024.19541 .

[4] Li, Luo, R. (2023). Improved Random Forest Based on Grid Search for Customer Satisfaction Prediction. *conomic Management and Green Development*, *1*(38), 198-207. https://doi.org/10.54254/2754-1169/38/20231913

[5] Ramin, H. (2024). Airline Customer Satisfaction. *1*. https://www.kaggle.com/datasets/raminhuseyn/airline-customer-satisfaction