DATA ANALYTICS AND APPLIED MATHEMATICS (DAAM)
e-ISSN: 2773-4854
VOL. 4, ISSUE 1, 7 – 22
DOI: https://doi.org/10.15282/daam.v4i1.9071

ORIGINAL ARTICLE

Prediction of US airline passenger satisfaction using machine learning algorithms

A.C.Y.Hong[1] , K.W. Khaw[1,*], X.Y. Chew[2], and W.C. Yeong[3]

[1] School of Management, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia
[2] School of Computer Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia
[3] School of Mathematical Sciences, Sunway University, Petaling Jaya, Malaysia

ABSTRACT – Due to the COVID-19 pandemic, the U.S. financial system and economy have also been severely affected. The U.S. airline industry has been hit particularly hard by the COVID-19 pandemic. Additionally, the aviation industry is also full of competition. One of the ways to attract customers and compete with other airline companies is by improving their service quality. Therefore, this study aims to predict the satisfaction of airlines based on the machine learning model and discover which features are more correlated with the target variable. In this study, the dataset consists of 129,880 observations and 1 target, 22 features or attributes (not including the identification). In this study, the result showed that the features that slightly correlate more with customer satisfaction are 'Online boarding', 'Inflight entertainment ', 'Seat comfort', 'On-board service', 'Leg room service', 'Cleanliness', 'Flight distance' and 'Inflight wifi service'. Then, K-Nearest Neighbors (KNN), Decision Tree Classifier (DTC), Logistic Regression (LR), Random Forest (RF), Naïve Bayes (NB) and AdaBoost were used to build the classification models. Data cleaning, exploratory data analysis, feature selection and One Hot Encoding were also performed before building the models. Finally, the models were evaluated based on their accuracy, precision, recall and F1-score. The results suggest that the champion model for this study is Random Forest, which achieved 89.20% accuracy, 93.04% precision and 88.80% F1-score. The results of this study can be used as a guide in applying machine learning to predict the satisfaction of airline passengers. This can also contribute to attracting passengers by improving the airline service quality.

ARTICLE HISTORY
Received: 06/02/2023
Revised: 28/02/2023
Accepted: 21/03/2023

KEYWORDS
Machine learning
Satisfaction prediction
K-nearest neighbors
Random forest
AdaBoost

INTRODUCTION

The aviation sector has grown to be one of the largest segments of the economy, allowing people and goods to move farther distances more quickly. However, due to the COVID-19 pandemic, the U.S. aviation sector has been particularly hard hit and is currently going through its harshest downturn. According to the Bureau of Transportation Statistics (BTS)[1], the number of passengers carried by U.S. scheduled service airlines during the calendar year 2021 was 670.4 million (preliminary), an increase of 303.6 million (83%) from 2020 and a decrease of 245.9 million (27%) from pre-pandemic 2019. Domestic flights carried 88% of all passengers in 2019, the final full calendar year before the start of COVID-19, while international flights carried 12% of all passengers. Domestic flights carried 91% of all passengers in 2021, while international flights carried 9%. After the pandemic, US airlines have also begun to resume some flights one after another, and the number of customers taking flights has gradually increased.

In this study, we will use some features related to service quality to predict passengers' satisfaction in US airlines. Most studies in the aviation sector in the past have been created to gauge various aspects of customer satisfaction and service quality [2-4]. According to the research, the result shows that airline service quality has a positive and significant relationship with passenger satisfaction [5].

This study aims to predict future passenger satisfaction using machine learning algorithms. The models are built using passenger satisfaction prediction data. The goal is to identify whether the passenger is satisfied or unsatisfied with US Airlines' service quality and to discover which features are correlated to passenger satisfaction. This study provides guidelines for airlines to provide adequate service levels to meet customer expectations and increase their competitive advantage. To improve customer satisfaction, several airlines have begun to focus on air service quality. Service quality conditions have an impact on a company's competitive advantage, so airlines should predict their future passenger satisfaction to improve their service quality and increase passenger loyalty.

The structure of this paper is as follows: a review of relevant literature, a description of the research methodology, the study's findings, and recommendations for future studies.

*CORRESPONDING AUTHOR | K.W. Khaw | ✉ khaiwah@usm.my

LITERATURE REVIEW

Customer satisfaction prediction

Customer satisfaction prediction is the process of estimating a customer's level of satisfaction based on previous data from satisfaction surveys. A subfield of artificial intelligence (AI) and computer science called "machine learning" aims to simulate human learning by using data and algorithms to gradually increase a system's accuracy. Machine learning includes two types: supervised machine learning and unsupervised machine learning. The labelled data is used in supervised machine learning to train algorithms that correctly identify data or predict outcomes. Otherwise, unsupervised machine learning analyses and groups of unlabeled datasets using machine learning techniques. This study utilized the supervised learning algorithms because the dataset was labelled and the values that are going to be predicted are “Satisfied” or “Neutral or dissatisfied”.

Several empirical studies have applied machine learning classifiers for customer satisfaction prediction on a real-world dataset. The study of airline passenger satisfaction is often measured using machine learning techniques and sentiment analysis, which looks at the text, tweets, or comments to determine positive or negative satisfaction [6-8]. In the absence of this, Leon and Martin [9] employ the Fuzzy segmentation approach to assess the technical quality and functional quality in the US airline sector to gauge airline passenger satisfaction.

In the research of Conlon et al. [10], they show that it is possible to analyse and accurately forecast employee job satisfaction using supervised machine learning approaches. According to the AUC measure, the Nystroem Kernel SVM Classifier algorithm performs the best, with an accuracy rate of more than 96%. The algorithm can also identify the most crucial elements that have a significant influence on forecasting work satisfaction. According to our research, a company's work culture and career opportunities both have a significant impact on forecasting employee job satisfaction.

619 of the 791 individuals (78.1%) expressed satisfaction with their psychotherapy sessions. The three most obvious factors that determine whether clients are satisfied with psychotherapy are the occupation of the clients, the location of psychotherapy, and the method of access to psychotherapy. With an F1 score of 0.758, the machine-learning model built on the CatBoost algorithm classified satisfied and psychotherapy clients with the maximum degree of accuracy [11].

In the research of Polce et al. [12], there were 413 patients in the study cohort, and 331 (82.6%) of them reported being satisfied two years after their surgery. For the independent testing set that was not used for model training, the support vector machine (SVM) model performed the best relative (c-statistic=0.80, Journal Pre-proof ML Satisfaction TSA calibration intercept=0.20, calibration slope=2.32, Brier score=0.11) model. Baseline single assessment numeric evaluation (SANE) score, exercise and activity, workers' compensation status, diagnosis, symptom duration prior to surgery, body mass index, age, smoking status, anatomic vs. reverse TSA, and diabetes were the most crucial variables for predicting satisfaction.

There have been some previous studies that used KNN to do predictions. According to the research, the k-Nearest Neighbor (KNN) technique was utilized in their study to measure camera tenants' customer satisfaction. For evaluating the accuracy of classification in data mining, the KNN algorithm-based classification method is excellent. The results show AUC (0.750), accuracy (98%), classification recall (86.67%), and classification precision (100%) [13].

In addition to the KNN algorithm being used to predict satisfaction, the decision tree can also be used to predict satisfaction. Here are the results of some recent work on decision trees. With model precision ratios ranging from 74.2 to 78.2%, the CART technique made it easier to categorize examples into homogenous groups by providing accurate classifications for all the decision trees (before, during, and after the flight). This outcome demonstrates the suitability of CART analysis as an approach for analyzing passenger satisfaction with airline services [14].

Talingting's research [15] found that predictive models such as Naïve Bayes, C4.5, and KNN algorithms were used. According to the simulation results, the C4.5 algorithm is the best model for predicting work satisfaction according to the perceptions of the 157 school administrators who participated in the survey from the Department of Education, Division of Surigao del Norte, Philippines. The accuracy of the C4.5, Naïve Bayes, and K-Nearest Neighbor (KNN) algorithms was 80.89%, 74.52%, and 71.97%, respectively.

Several machines learning models, including Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Decision Tree classifiers, and Multi-layer Perceptron classifiers (MLP), are created for predicting parental satisfaction levels in the study Lal Mukherjee and Dutta [16]. The Decision Tree model has the lowest accuracy, according to a comparative study, at 82.29%. In Poushy et al. [17] analysis, they obtained 93.75% (LR and SVM) accuracy in their dataset, which indicated that the majority of students are dissatisfied with the way that online classes are currently being run and the speed of the internet.

There have been some previous studies about Naïve Bayes algorithms. The preparedness, empathy, reliability, and responsibility features of the Naïve Bayes approach were found to provide test results with an accuracy level of 85.48%, a precision value of 81.08%, and a recall value of 93.75% in the study by Aisyah et al. [18]. Based on the test results, the Naïve Bayes Method can be used to predict how satisfied a lecturer will be with the performance of the institution. In Poushy et al. [17] analysis, they obtained 93.75% (LR and SVM) accuracy in their dataset, which indicated that the majority of students are dissatisfied with the way that online classes are currently being run and the speed of the internet. In contrast, when compared to LR, KNN, and SVM, Naïve Bayes is the classifier with the lowest accuracy. It only receives 85%, which is below the 90%. In comparison to other models and algorithms used in WEKA, Naïve Bayes has the lowest accuracy for classifying the customer satisfaction level in the study by Roy et al. [19].

According to the study of Li et al. [20], the Adaboost algorithm can be utilized to forecast follow-up and control satisfaction data for diabetics. The final comparative study findings showed that the Adaboost algorithm performed the best for the testing dataset, with an accuracy of 94.84% and sensitivity and specificity of 95.76% and 93.56%, respectively. The experimental results showed that the Adaboost algorithm performed optimally in four models for the testing dataset, with an AUC of 0.9817 and a G-mean of 0.9465. The Adaboost algorithm may be successfully used for the health management control satisfaction of diabetes patients, according to the results presented in the research. According to the study by Bouzakraoui et al. [21], AdaBoost had the best accuracy, at 98.66%. It demonstrates how it can gauge consumer satisfaction by analyzing facial expressions.

It has several uses in daily life, including image classifiers, recommender systems, and feature pickers. Its real-world applications include disease prediction, loan default prediction, and fraud detection. For example, doctors can use the disease dataset to predict whether the patient is affected by that disease or not. To determine whether a patient has diabetes, heart disease, or cancer, artificial intelligence has classified several disease datasets using the Naïve Bayes classification and Random Forest algorithms. Hence, these data are only used by a doctor for analysis, who uses them to appropriately assess the patient's health status using his or her medical expertise [22]. Moreover, algorithms are also used in predicting loan default. The sustainable and healthy growth of P2P online lending platforms is associated with a rise in the probability of user loan default. As a result, using actual user loan data from Lending Club, their study creates a loan default prediction model based on the Random Forest algorithm. The experimental findings demonstrate that the Random Forest method performs better in predicting default samples than Logistic Regression, Decision Tree, and Support Vector Machine, and has a great ability to generalise [23].

The algorithms can also apply in the banking industry to detect fraud effectively. This is because the number of fraudulent activities is rising. Thus, both cardholders and the institutions that issue the cards need to adopt a methodical fraud detection system. The study of Sharma et al. [24] illustrated how several machine learning algorithms may effectively detect fraud. As a result, they recommend using Random Forest and ANN as the algorithms of choice when predicting the performance of credit card fraud detection systems. Similarly, the algorithms can use in the aviation industry. This study discussed US airline customer satisfaction, but not only US airlines can refer to this study, but other airline companies can also refer to this study to improve their service quality. Hence, the algorithms are useful and can address some issues effectively in the real world.

METHODOLOGY

In this section, the concepts of machine learning methods and the process to run the data have been discussed.

Technical background of the machine learning methods

K-nearest neighbors (KNN)

Due to its simplicity, efficiency, and intuitiveness, the k-Nearest Neighbor (kNN) classifier is one of the most well-known classification methods. It is a supervised classification model where the nearest neighbour number as well as the distance metric can be changed [25]. It is a non-parametric model that can be used for both classification and regression, which means it does not rely solely on feature similarity to infer anything about the data it is working with [26], [27]. The reason it is referred to as a lazy learner algorithm is that it saves the dataset rather than instantly starting to learn from the training set and then applying an action to it when it comes time to categorize it. It then calculates a pairwise distance or similarity measure for each training instance and each unseen example, selecting the k instances that are most similar to the unseen case [28], [29]. It selects k neighbours first and then calculates the Euclidean distance between them [30]. The formula below is the formula of Euclidean distance.

$$\mathrm{d} = \sqrt{(x_2 - x_1)^2 + (\mathrm{y}_2 - \mathrm{y}_1)^2} \tag{1}$$

Decision tree classifier (DTC)

A normal tree has leaves, branches, and a root system. The format is the same for the Decision Tree. It has branches, a root node, and leaf nodes. Each internal node has an attribute test, which returns a result on the branch, a class label, and a leaf node [31], [32]. In a decision tree, each node corresponds to an attribute, each link to a rule, and each leaf to the outcome (categorical or continuous value) [32].

A supervised learning approach known as a decision tree can be used to address problems involving classification and regression. Decision tree learning is one of the most well-liked and effective methods for inductive inference over supervised data. A decision tree depicts a method for categorizing categorical data based on several attributes. Decision trees can process huge amounts of data, which makes them valuable in data mining applications as well. Building decision trees does not require any parameter setup or domain knowledge. As a result, decision trees are appropriate and ideal for gathering information through exploratory analysis, and their representation of collected knowledge in tree form is natural and easy to understand [33].

The ID3, C4.5, and CART decision tree algorithms are the most often utilized ones. The criteria with the highest information gain are divided using ID3 to maximize information gain. While CART chooses the splitting attributes using the Gini Index, C4.5 is an extension of the ID3 algorithm and chooses the attribute with the highest gain ratio [33].

Logistic regression (LR)

The simplest machine learning approach and one of the generalized linear models (GLM) is logistic regression. The LR is a popular supervised machine learning (ML) binary classification technique. Compared to standard linear regression, the logistic regression model is more robust [34], [35]. It can be applied to categorical dependent variables, with binary or discrete categorical variables 0 or 1 as the outcome. Hoffman [36] describes in their study the Multiple Explanatory Variables by extension of the basic principles. The general equation is

$$\mathrm{P} = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i x_i)}} \tag{2}$$

where

- X is the input value.
- P is the probability of success.
- β_0 is the bias or intercept term.
- β_I is the coefficient of the X variate.

Additionally, are the model's parameters, e is the mathematical constant known as Euler's number, which is approximately equivalent to 2.78. Multiple independent variables, which can be continuous or categorical, can be used in logistic regression. Logistic regression is comparatively speedy but still accurate enough when compared to other classification methods. Since both methods are overly generic for complex relationships between variables, the issue is also present with linear regression [3].

Random forest (RF)

A supervised learning technique called the Random Forest classifier can be applied to classification and regression problems [37]. An expansion of decision trees is Random Forest. RFTs outperform decision trees, bootstrapping, and bagging, as is well known [38]. Growing an ensemble of trees and allowing them to select their favourite class has significantly improved classification accuracy. These ensembles, which control the growth of each ensemble tree, are commonly grown using random vectors. An early example is bagging, in which each tree is grown using a random pick (without replacement) from the training set's examples [39]. An RF needs to be trained using two parameters, the number of trees in the forest (ntree), and the number of randomly selected features or variables that are used to evaluate each tree node (mtry). The voting cutoff, which can be altered in RF, is used to determine recall, accuracy, and F1 score.

Naïve Bayes (NB)

The Naïve Bayes Classifier, one of the most well-known supervised machine learning techniques, is a simple probabilistic classifier built on the Bayes theorem. Each attribute is independent when Bayes' theorem and "Naïve" are combined [40]. The assumption of attribute independence, which may be broken in many real-world data sets, is what gives it its efficiency. The assumption has been addressed in a variety of ways, with attribute selection being one of the most crucial. Traditional approaches to attribute selection in Naïve Bayes, on the other hand, have a high computational overhead [41].

This approach has several significant benefits, one of which is that it only needs a minimal dataset for training. This approach allows for the estimation of classification-related parameters and seeks to determine the posterior probability [42], [43]. A mathematical formula for calculating the likelihood of a hypothesis given available evidence is the Bayes theorem, commonly known as Bayes' Rule or Bayes' law. The deciding element is conditional probability. The posterior probability, $P(c|x)$, can be calculated using the Naïve Bayes theorem method using $P(c)$, $P(x)$, and $P(x|c)$. The formula below can be used to get the posterior probability.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \tag{3}$$

where

- P(c|x) is the posterior probability of class (target) given predictor (attribute).
- P(c) is the class prior probability.
- P(x|c) is the likelihood which is the probability of the predictor given class.
- P(x) is the predictor prior probability.

In Naïve Bayes classification, the initial step is to create a frequency table for each characteristic against the target. The next step is to build likelihood tables by finding the probability of the provided features. Finally, calculate the posterior probability for each class using the Naïve Bayes equation. The class with the highest posterior probability is the outcome of the prediction [44]. The Naïve Bayes classifier has demonstrated success in a variety of applications, including text categorization and medical diagnosis.

AdaBoost

It is currently common to look for ensemble learning approaches to improve categorization accuracy. From this vantage point, the classifiers have been significantly enhanced by the introduction of ensemble approaches. To iteratively categorize the data set and provide a reliable classification, these techniques combine several weak learners or classifiers. Boosting algorithms are the most popular ensemble techniques [45]. In the research of Schapire and Singer [46], Adaptive Boosting is one of these methods and is a variation of the Boosting algorithm (AdaBoost).

The main idea behind AdaBoost is to repeatedly apply the same weak learning algorithm, W, to the training data in different probability distributions. It starts with a uniform distribution. It assigns greater weights to samples that the previous classifier incorrectly classified to help the current weak classifier concentrate on those examples. Samples that were simultaneously correctly recognized will also receive lower weights. The W distribution is then altered after each cycle. Ultimately, a single "strong" hypothesis is produced from the hypotheses produced by each cycle's weak learner.

AdaBoost's outstanding performance is due to its capacity to maximize headroom on a training set, which enhances classifier performance. It should be mentioned that Boosting has proven to be quite successful in resolving two-class classification issues. We carry out numerous binary classifications to get a multi-class classification.

Flow chart

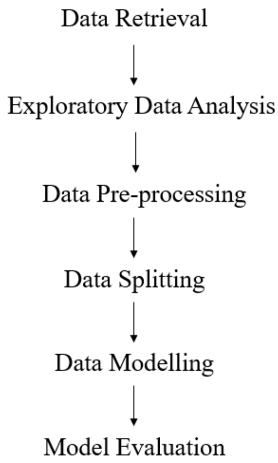

Figure 1. Flow chart of the overall methodology

Data retrieval

Airline Passenger Satisfaction Data, the dataset utilized for this study, offers details regarding the quality of service provided by US Airlines. It is freely accessible and is available for download from Kaggle. The read csv () function from Pandas was used to import the data into Python once it was downloaded in the CSV file format. With the help of the read csv () function, a CSV file can be converted into a data frame or a 2-dimensional labelled data structure with columns of various types. One customer's satisfaction with various aspects of service quality is represented by each row. The dataset has 129,880 observations and 1 target, 22 features or attributes (not including the identification). There are four categorical data in the dataset, which are ‘Gender’, ‘Customer Type’, ‘Type of Travel’, ‘Class’. Satisfaction is the target variable or the column that is going to be predicted.

Exploratory data analysis (EDA)

Descriptive statistics

Regarding distribution, central tendency, and variability for continuous variables as well as the proportion for categorical variables, descriptive statistics are used to summarize the characteristics of the features in the dataset. Using the describe() function of Pandas, the descriptive analysis for the continuous variables was carried out. When it comes to categorical variables, some proportions were depicted using a Matplotlib pie chart (pie()) with the option to label the percentage and others using a Seaborn factor plot (factorplot()) to draw a categorical plot onto a FacetGrid [47].

Correlation between features

Only the linear correlation between two variables is revealed by the correlation heatmap. It suggests that there might be a direct connection between them. Additionally, this plot makes it simple to analyze the dataset and comprehend how variables relate to one another. Seaborn heatmap (heatmap ()) is used to plot the correlation matrix.

Correlation with the target variable

In this part, correlation (corr()) and horizontal bar chart (‘barh’) are used to describe the correlation with the target variable (satisfaction). In this part, service scoring features and age features are used to do the correlation with satisfaction after deleting the ‘Arrival Delay in Minutes’ and ‘Departure Delay in Minutes’ variables.

Data pre-processing

After the data was collected, data pre-processing was done to put the data into a format that the prediction models could identify. Four subtasks make up the data pre-processing for this project: deleting unnecessary columns, finding missing values, feature selection and encoding categorical variables.

Deleting unnecessary columns

The original dataset includes the 'Unnamed: 0' and 'id' column, which is not used in this study. Therefore, by using Pandas' dataframe.drop () function, this column was eliminated from the data frame. After Exploratory Data Analysis, the 'Departure delay in minutes' and 'Arrival delay in minutes' columns have dropped. After the deletion, there were only 21 columns left including the target column.

Finding missing values

The Pandas function dataframe.isnull ().sum() was used to return the number of missing values in each of the dataset's columns to determine whether any values were missing. Only 'Arrival Delay in Minutes' has 393 missing values. The rows which contain nulls were simply removed because they only made up less than 1% of the data which is 99.70% of the filling factor. When the dataset is huge, the common technique used to handle missing values is to remove them. Since the 393 rows are only less than 1% of the data, removing these rows would not likely lead to a loss of information that would alter the result. Besides, there was still a large enough sample in the dataset.

Replace the target values

The function replace () from Pandas was used to replace the 'neutral or dissatisfied' with 0 and replace the 'satisfied' with 1.

Feature selection

Feature selections are used to select the features that are important to the target variable. In this study, only 5 of 20 features are selected as feature importance to predict the target variable. To get the important features, we used the coefficient (model.coef_[ ]) function of the logistic regression classifier to do the selection.

Encoding categorical variables by using One Hot Encoding

The original dataset consists of 4 categorical variables which are 'Gender', 'Customer Type', 'Type of Travel' and 'Class'. These 4 columns contain textual data. Meanwhile, machine learning algorithms can only take numerical input and output. Therefore, these textual or categorical data must be encoded as integer values before they can be used to train and test the models. In this study, One Hot Encoder from Scikit-learn pre-processing library was used to encode the categorical variables into numerical variables. All the categorical values were transformed into a value that is 0 and 1. After selection features, only 'Gender', 'Customer Type' and 'Class' were encoded using OneHotEncoder() function.

Data splitting

The study demonstrates that data splitting, in which a given dataset is split into two distinct sets for training and testing, is a common method for model validation. The statistical and machine learning models are next fitted to the training set, and they are eventually tested on the testing set. By providing a set of data for validation apart from training, we may assess the effectiveness of various models without considering any bias that was introduced during their training [48].

The dataset was divided into a 25% test set and a 75% training set for this investigation. The Scikit-learn model selection library's train_test_split function was used for this. This function with a specified parameter extracts 75% of the samples together with the corresponding label as the training set and the remaining 25% with the corresponding label as the test set. The parameter random_state was also fixed so that the same output will be produced every time the function was run.

Data modelling

There have six classification models (KNN, DTC, RF, LG, NB, AdaBoost) were built at this stage. The algorithm and the provided parameters are initially stored in an object for each of the models. The data that the algorithm pulled from the training set is likewise contained in this object. Then, applying the fit method to the earlier-created object to take the argument as NumPy arrays, *X_train* and *y_train*, where X is the features and y is the target, the classification model was formed on the training data. The prediction can then be produced by executing the prediction method on the object that contains the test set's features (*X_test*).

The parameters utilised for each of the classifiers in this study are further described in the section that follows.

K-nearest neighbors (KNN)

KNeighborsClassifier() from Scikit-learn was used to create this classifier. The range from 1 to 25 was tested to find the value of k that provides the maximum accuracy. The findings revealed that $k = 19$ produced the best accuracy. Thus, the parameter of the given function was set to n_neighbors = 19.

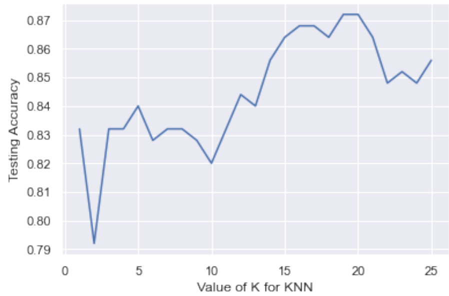

Figure 2. Finding the best value of k

Decision tree classifier (DTC)

Using the DecisionTreeClassifier() method from Scikit-learn, this classifier was created. The default parameters were used because no parameters were supplied. By default, the minimum number of samples required to split an internal node was set to 2 (min_samples_split=2), and this function uses the Gini impurity to assess the split's quality (criterion="gini").

Logistic regression (LR)

Using the LogisticRegressionClassifier() function from Scikit-learn, this classifier was created. The parameter was used that is, random_state also set as 42. The random number generator was used to get consistent values.

Random forest (RF)

RandomForestClassifier() function from Scikit-learn was used to create this classifier. The n_estimators option controls the number of trees in the classifier. In Scikit-learn, the default value of n_estimators is 100. The random_state was set as 42.

Naïve Bayes (NB)

This classifier was created using Scikit-learns GaussianNB() function. For this function, no parameters were supplied. The default parameters of this function are adjusting the prior probabilities according to the data (priors=None) and setting the portion of the largest variance of all features to 1×10^{-9} (var_smoothing=1e-09). This parameter is added to the variance to stabilize the calculation [49].

AdaBoost

Using the AdaBoostClassifier() function from Scikit-learn, this classifier was created. The default parameters were used because no parameters were supplied.

Model evaluation

Four performance metrics were used to quantify the performance of the classifiers: Accuracy, Precision, Recall and F1-score. The performance metrics are based on the confusion matrix in Table 1. The confusion matrix is a particular table design that enables visualizations of a classifier's performance. Large True Positive and True Negative numbers and tiny False Positive and False Negative values are characteristics of a strong classifier.

Table 1. Confusion matrix

		Predicted Class	
		Purchase (1)	**Does Not Purchase (0)**
Actual Condition	**Purchase (1)**	True Positive (TP)	False negative (FN)
	Does Not Purchase (0)	False Positive (FP)	True Negative (TN)

Accuracy

An indicator of how accurately a model has categorized the records is the accuracy or total success rate. It is calculated by dividing the total number of instances by the sum of TP and TN. It measures the percentage of predictions that were made correctly. The misclassification rate, which measures the proportion of results that were incorrectly anticipated, is also frequently present. A high accuracy score alone does not indicate that the model is well-established; a solid model should also have a decreased misclassification rate. The formulas are shown below in (4) and (5).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{TP + FP + FN + TN} \tag{4}$$

$$\text{Misclassification} = 1 - \text{Accuracy} \tag{5}$$

Due to the unbalanced dataset, it can be misleading to choose the best-performing model just based on accuracy and misclassification rate. By accurately anticipating the majority or positive class, high accuracy and a low misclassification rate can be attained in this situation. As a result, the champion model was also derived using the other measures, which are described in the following subsections.

Precision

Precision is the percentage of instances marked positive that are positive. The formula of precision is shown below in (6).

$$\text{Precision} = \frac{\text{TP}}{TP + FP} \tag{6}$$

Recall

Recall or sensitivity is the percentage of positive instances that were correctly identified. The formula recall is shown below in (7).

$$\text{Recall} = \frac{\text{TP}}{TP + FN} \tag{7}$$

F1-score

The F Score or the F Measure are other names for the F1-score. The harmonic mean of recall and precision when both measurements are taken into account is the F1 score. The balance between recall and precision is represented by the F1 score. The formula of the F1 score is shown below in (8). When working on classification models with an unbalanced data set, the F1 score is extremely useful. Similarly, the higher F1-score indicates the classifier is performing well as it can predict most of the positive observations.

$$\text{F1} - \text{score} = 2 \times \frac{Precision \times Recall}{Precsion + Recall} \tag{8}$$

EMPIRICAL RESULTS

Exploratory data analysis (EDA)

Visualizing categorical features

Target variable: satisfaction

Table 2. Number and percentage of satisfaction column

Satisfaction	Number	Percentage
Satisfied	56,428	43.45%
Neutral or dissatisfied	73,452	56.55%

The target column consists of two categories which are “satisfied” and “neutral or dissatisfied”. The amount of neutral or dissatisfied passengers is higher than the number of satisfied passengers in this data. As shown in Table 2, this data does not have any imbalance problem.

Customer information

Table 3. Satisfaction based on gender

	Satisfaction (Female)	Satisfaction (Male)
Satisfied	42.90%	44.01%
Neutral or dissatisfied	57.10%	55.99%

Table 3 shows that the satisfaction rates of women and men, both are around 43-44%. There is no dominance in satisfaction by gender. The dissatisfaction rate is higher in both genders.

Table 4. Satisfaction is based on age group

Age group	Satisfaction	
	Neutral or dissatisfied	Satisfied
[7-27]	17.0%	6.9%
[27-40]	15.2%	10.2%
[40-51]	10.5%	14.3%
[51-86]	13.9%	12.0%

As shown in Table 4, while the majority of passengers between the ages of [40,51] are satisfied, the rate of dissatisfaction is higher for passengers in other age ranges.

Table 5. Satisfaction probability based on customer information

	Customer Type		Type of Travel		Class		
Satisfaction probability	Loyal Customer	Disloyal Customer	Business Travel	Personal Travel	Business	Eco Plus	Eco
	0.478115	0.239697	0.583724	0.101326	0.694434	0.246414	0.187673

In the Customer Type feature, which is divided into two groups Loyal customers and disloyal customers, the probability of Loyal customers is more than the probability of Disloyal customers. It means Loyal customers most satisfied than the disloyal customers. Furthermore, the Type of Travel feature consists of two categories as Personal and Business travel. It seems that the probability of passengers making Business travel is higher than those making Personal travel. It means 58% of passengers travelling on business are satisfied, and 90% of passengers travelling on personal are neutral or dissatisfied with the flight. Class features are divided into three categories: Eco, Business and Eco Plus. While 69% of passengers class on business are satisfied, 25% of passengers class on Eco plus are satisfied and 19% of passenger class on eco are satisfied.

Service scoring

Table 6. Average satisfaction and standard deviation of service ratings

Features	Average satisfaction	Standard deviation
Inflight service	**3.6**	1.2
Baggage handling	**3.6**	1.2
Checkin service	3.3	1.3
Gate location	3.0	1.3
On-board service	3.4	1.3
Cleanliness	3.3	1.3
Leg room service	3.4	1.3
Seat comfort	3.4	1.3
Inflight wifi service	**2.7**	1.3
Food and drink	3.2	1.3
Inflight entertainment	3.4	1.3
Online boarding	3.3	1.4
Ease of Online booking	2.8	1.4
Departure/Arrival time convenient	3.1	1.5

As shown in Table 6, the features with the highest average satisfaction rate are Inflight_services and Baggage_handling with an average of 3.6. The feature with the lowest satisfaction rate is Inflight_wifi_service with an average of 2.7. Table 6 also shows the standard deviation to account for any deviations between ratings. They are close to each other.

Visualizing numeric features

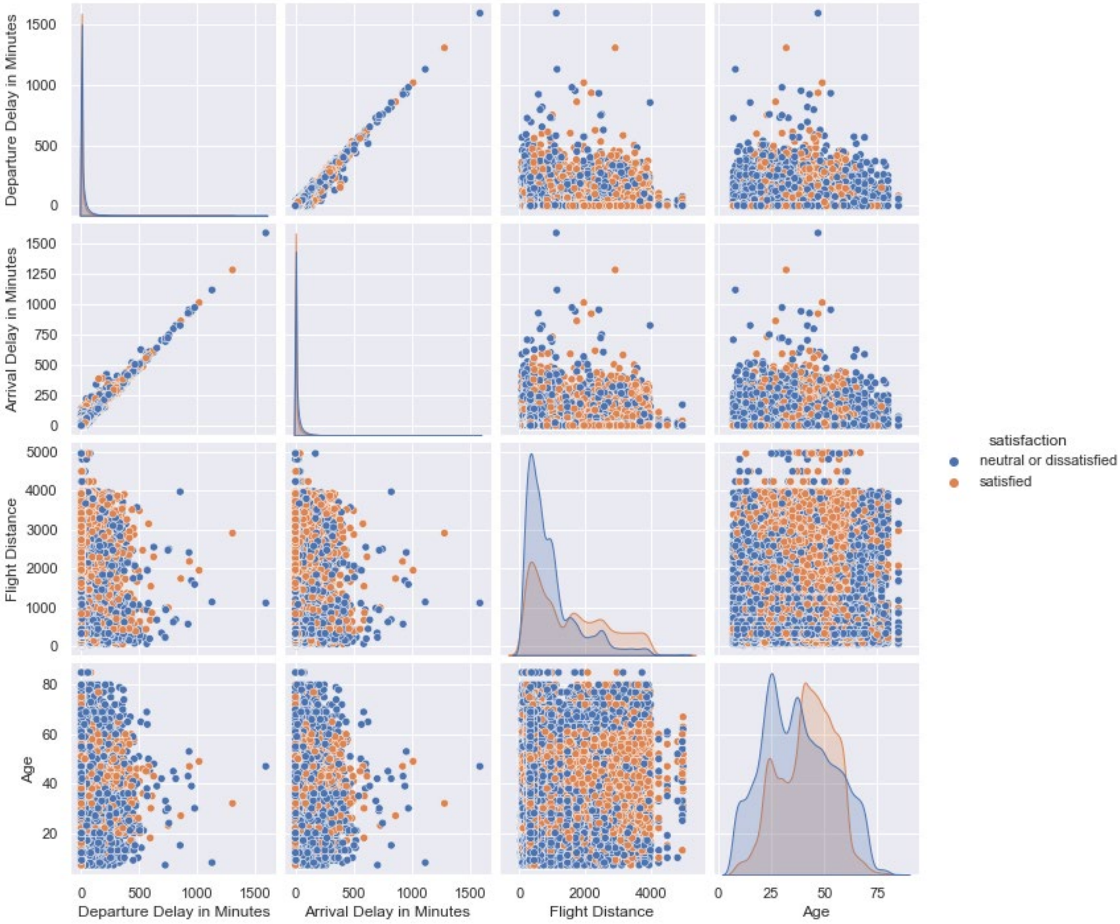

Figure 3. Pairplot about the numeric features

While there is a visible relationship between some numeric features (Arrival_Delay_in_Minutes and Departure_Delay_in_Minute), some are unrelated to each other (Flight_Distance and Age).

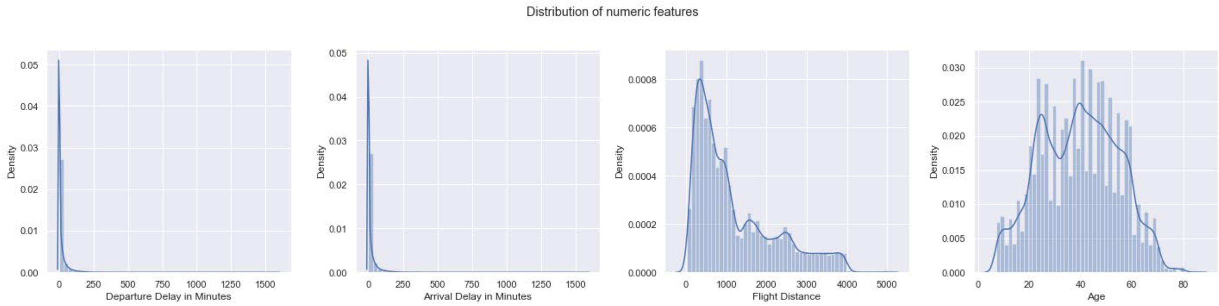

Figure 4. Subplots about the distribution of numeric features

The maximum value for the Arrival Delay in Minutes and Departure Delay in Minutes columns is 0 Occurrences decline as the delay minutes rise. The numbers in the Flight distance column are generally confined to the 0-1000 range. Additionally, the data includes individuals of every age. Moreover, Arrival_Delay_in_Minutes and Departure_Delay_in_Minutes columns are highly positively correlated. Correlated features will be checked again with a heatmap.

Correlation between features

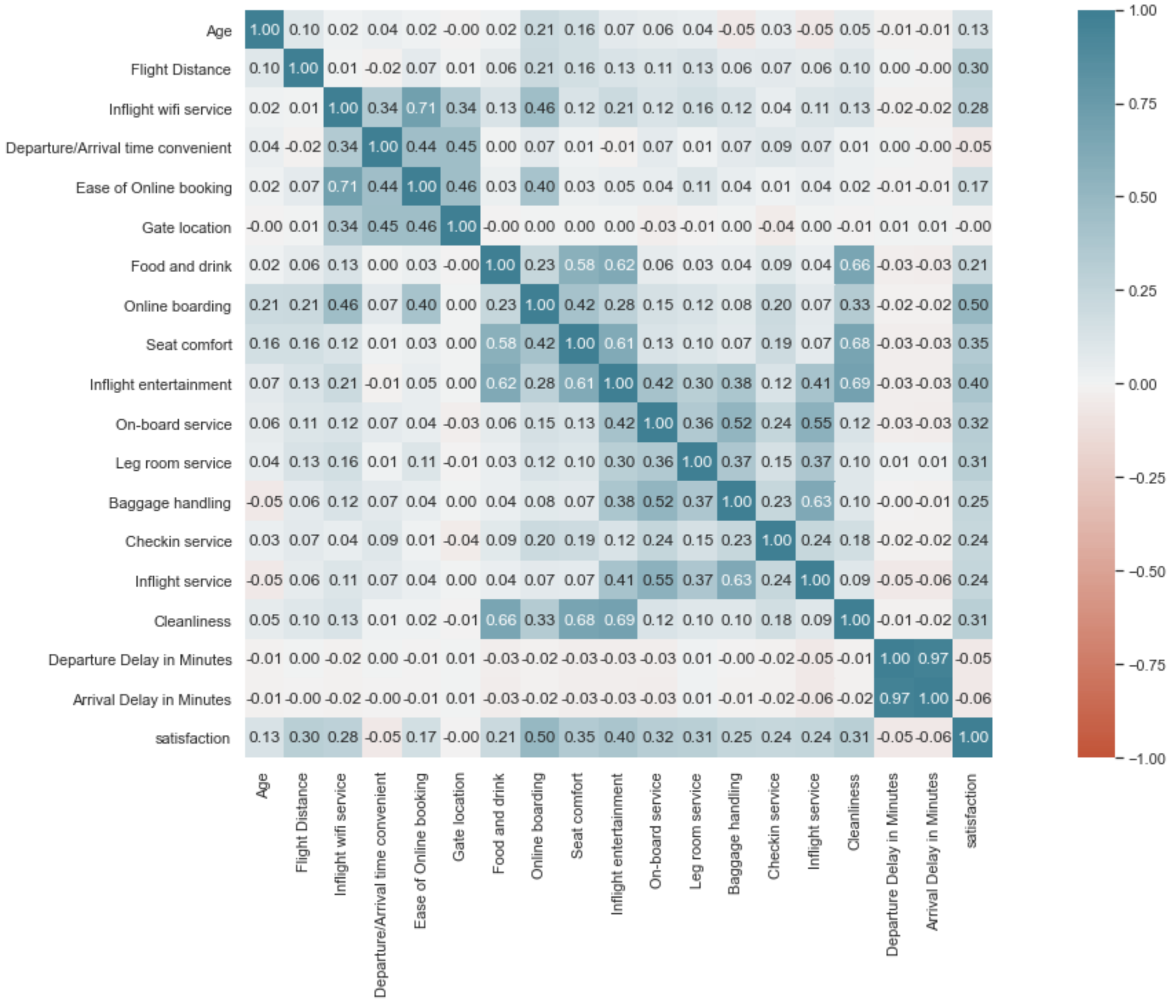

Figure 5. Heatmap about the correlation between features

As shown in the heatmap, "Ease of Online booking" and "Inflight wifi service" are positively correlated with a ratio of 0.71. "Cleanliness" and "Inflight entertainment" are positively correlated with a ratio of 0.69. "Cleanliness" and "Seat comfort" are positively correlated with a ratio of 0.68. "Cleanliness" and "Food and drink" are positively correlated with a ratio of 0.66. "Inflight service" and "Baggage handling" are positively correlated with a ratio of 0.63. "Inflight entertainment" and "Food and drink" are positively correlated with a ratio of 0.62. Also, "Inflight entertainment" and "Seat comfort" are positively correlated with a ratio of 0.61.

The heatmap reveals that the "Departure delay in minutes" and "Arrival delay in minutes" columns have a strong positive correlation (0.97). The 'Arrival_Delay_in_Minutes' column has null values. Additionally, the columns labelled "Departure delay in minutes" and "Arrival Delay in Minutes" are filled with zero values, indicating that they are not key components of the model. Both columns will drop.

Correlation with target

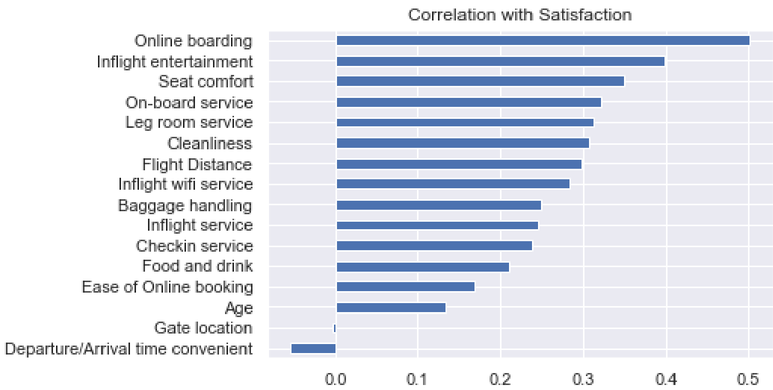

Figure 6. Correlation between target variables

According to Figure 6 above, features that slightly correlate more with customer satisfaction are 'Online boarding', 'Inflight entertainment', 'Seat comfort', 'On-board service', 'Leg room service', 'Cleanliness', 'Flight distance' and 'Inflight wifi service'. Among features "Online boarding" has the maximum correlation to the target, we will check its correlation with other features.

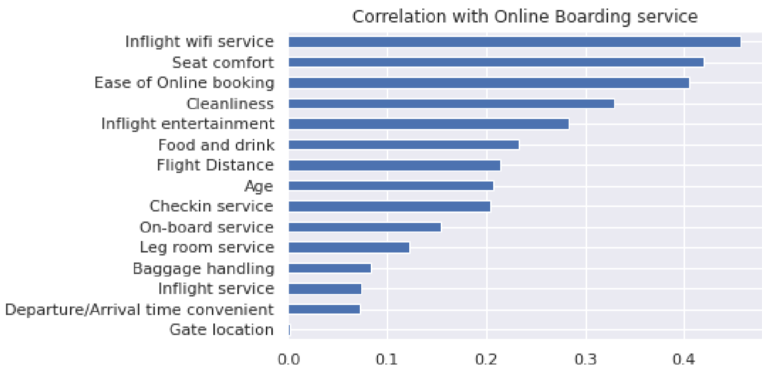

Figure 7. Correlation with Online boarding service

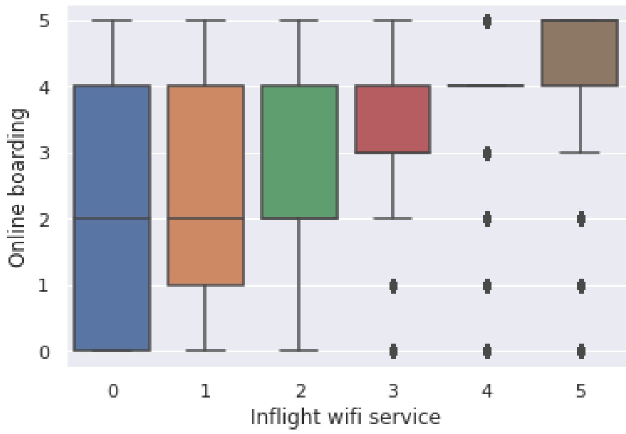

Figure 8. Boxplot about the correlation between Inflight wifi service and Online boarding

As the score given to the Inflight wifi service increases, the range distributed to Online boarding decreases and its score increases. People who get better service of Inflight wifi are more likely to give better ratings for Online boarding.

Performance results

Before data modelling and evaluation, this study had done the feature selection. In this study, only 5 of 20 features are selected as feature importance to predict the target variable. These are the 5 features' importance: 'Customer Type', 'Departure/Arrival time convenient', 'Online boarding', 'Inflight entertainment' and 'Cleanliness'.

Comparison of accuracy

In this study, six supervised machine learning algorithms are used to predict satisfaction. After running the Python, it shows that the accuracy of six algorithms which are Random Forest (89.2%), K-Nearest Neighbors (87.2%), Decision Tree Classifier (82.00%), AdaBoost (82.8%), Logistic Regression (78.4%) and Naïve Bayes (76.8%). It shows that **Random Forest is the highest accuracy** compared with the other five models which are 89.2% and the **Naïve Bayes is the lowest accuracy model**.

Comparison of confusion matrix

When comparing the confusion matrix of six models, the results show that **Random Forest** outperforms the other five models by predicting more true values and fewer false values. Random Forest predicts a total of 223 true values and 27 false values.

Comparison of precision

In this study, the precision of six models achieved similar performance in terms of precision. This is because all the models achieved a precision of 0.83-0.93 or 83%-93%. The highest precision value is **Random Forest** which is 93.04% while **Decision Tree Classifier** scored the lowest value which is 83.47%.

Comparison of recall

After comparing the results of recall, **KNN** achieved the highest score in terms of recall which is 88.10% and **Naïve Bayes** is the lowest recall value of the six models. The Naïve Bayes is the only model was achieved a score lower with only 61.90%. The Logistic Regression also achieved slightly lower with only 69.05% which is below the 70%.

Comparison of F1-score

In this study, KNN, Random Forest, Decision Tree and AdaBoost classifiers achieved similar performance in terms of F1-score, as all four classifiers achieved an F1-score of 0.82-0.88 or 82%-88%, with the highest F1 achieved by **Random Forest (0.8880)**, followed by KNN (0.8740), AdaBoost (0.8259) and DTC (0.8178). Meanwhile, **Logistic Regression (0.7632)** and **Naïve Bayes (0.7290)** scored slightly lower than 80%.

Best model: random forest (RF)

After comparing the average score of six classifiers with the help of assessment tools such as the accuracy, precision, recall and F1-score, **Random Forest (88.99%)** outperforms the other five models in terms of **accuracy, precision and F1-score**. It is followed closely by K-Nearest Neighbor (KNN) with an 87.36% average score. The models achieved about 80% average score, which is AdaBoost (82.66%) and Decision Tree (81.85%). Logistic Regression (77.27%) and Naïve Bayes (75.06%) achieved a lower average score that is below 80%. Therefore, it can be concluded that **Random Forest** is the classifier that must be used in predicting the satisfaction of airline passengers instead of using K-Nearest Neighbors, Decision Tree, AdaBoost, Logistic Regression and Naïve Bayes.

Table 7. Summary of performance outcome (%)

	KNN	DTC	LR	RF	NB	AdaBoost
Accuracy	87.20	82.00	78.40	**89.20**	76.80	82.80
Precision	86.72	83.47	85.29	**93.04**	88.64	84.30
Recall	**88.10**	80.16	69.05	84.92	61.90	80.95
F1-score	87.40	81.78	76.32	**88.80**	72.90	82.59
Average	*87.36*	*81.85*	*77.27*	***88.99***	*75.06*	*82.66*

DISCUSSION

This chapter will discuss the results of EDA and discuss accuracy, precision, recall, F1-score, as well as the best model to predict satisfaction. After creating the heatmap, there is no multicollinearity and some features that slightly correlates more with passenger satisfaction. According to Hulliyah [50], their study employed several classification models, including KNN, Logistic Regression, Gaussian NB, Decision Trees, and Random Forest, to predict airline passenger satisfaction. The Random Forest Algorithm can reach an accuracy of 99% with a threshold of 0.7, and the Inflight Wi-Fi Service is essential for attaining customer satisfaction, according to the study's findings. Compared with a similar study, this study found out Random Forest Algorithm achieved an accuracy of 89.20% even though the accuracy is lower than Hulliyah's study. In addition, this study also found out the 'Online boarding', 'Inflight entertainment', 'Seat comfort', 'On-board service', 'Leg room service', 'Cleanliness', 'Flight distance' and 'Inflight wifi service' are more correlated with passenger satisfaction. This result had more features that correlate with passenger satisfaction and it can help the airline company improve their service quality. According to Tsafarakis et al. [51], 'Inflight entertainment' enhancement and 'Inflight wi-fi service' that caters to customers' needs can enhance airline passenger satisfaction from the flight criterion. In order to increase customer satisfaction and the likelihood that they would return, Park [52] suggests that online airline services and in-flight experiences be upgraded.

After comparing the average score of six classifiers with the help of assessment tools such as the accuracy, precision, recall and F1-score, Random Forest (88.99%) outperforms the other five models in terms of accuracy, precision and F1-score. It is similar to the study of Gao et al. [53], whose result shows that Random Forest is *the highest accuracy (95.92%)* compared with Logistic Regression, Naïve Bayes, Support Vector Machine and MLP in modelling airline travel satisfaction. In the study of Jiang et al. [54], the classification prediction model with the best classification performance uses Random Forest following RF-RFE feature selection. This feature subset's RF has the most effective classification performance (accuracy: 0.963, precision: 0.973, recall: 0.942, F1 value: 0.957, AUC value: 0.961). Additionally, several classifier models were applied to the dataset, and Random Forest (RF) performance takes first place with the highest accuracy rate (91.53%) for predicting Bangladesh's internet user satisfaction levels [55]. In the research of Shetu et al. [56], the results show that when predicting customer satisfaction with Bangladesh's online banking system, three algorithms—KNN, Logistic Regression, and Random Forest achieved the best accuracy of 96%. Precisions are 98%, recall is 96%, and F1-score is 97% for these three models. After comparing with other studies, their Random Forest achieved the highest accuracy and their accuracy was higher than that of this study, but the Random Forest of this study also achieved the highest accuracy. In contrast, the findings of the study by Jain et al. [57], the Random Forest approach is the most accurate algorithm for predicting employee satisfaction for Om Logistics, with an accuracy of 87.3% and a sensitivity of 0.89. The accuracy of Random Forest is lower than this study when Random Forest of this study scored an accuracy of 89.20% and it achieved the highest accuracy. The results of them are consistent with what we found. Hence, Random Forest is the best classification model to do the prediction of passenger satisfaction in this study.

CONCLUSION

In conclusion, the objectives of this study were achieved. This study aims to predict the satisfaction of airlines using machine learning. Four performance metrics were used to evaluate the performance of the classifiers which are accuracy, precision, recall and F1-score. Next, each of the six chosen classifiers was trained with the dataset using a 75:25 train-to-test ratio and was evaluated based on the four performance metrics. *Random Forest (RF)* was the best-performing model, followed by KNN, AdaBoost, Decision Tree Classifier, Logistic Regression and Naïve Bayes.

Additionally, the features that slightly correlate more with customer satisfaction are 'Online boarding', 'Inflight entertainment', 'Seat comfort', 'On-board service', 'Leg room service', 'Cleanliness', 'Flight distance' and 'Inflight wifi service'. After that, we selected 5 of 20 features as being important to prediction using feature importance with logistic regression classifier. These are the 5 features' importance: 'Customer Type', 'Departure/Arrival time convenient', 'Online boarding', 'Inflight entertainment' and 'Cleanliness'. The purpose of predicting airline passenger satisfaction is to help US Airlines' passengers increase their satisfaction, specifically by improving service quality.

An important implication of our research is that it gives airline companies a clear understanding of the numerous alternatives they have, as well as a faster and more efficient way to predict passenger satisfaction with higher accuracy. The contribution of this study is that it may be used as a reference for these companies as they choose which machine learning algorithm will have the highest probability of accurately predicting consumer satisfaction.

This research was limited by the fact that not all necessary factors were included in the dataset and that only 5 important features were selected to predict satisfaction. Additionally, the study was conducted using four performance metrics only, which may make the results less detailed and in-depth. Finally, the findings of this study are restricted to the data contained in the defined dataset and cannot be used to make generalizations about other datasets. Future research can expand the scope of the study by considering the significance of each feature in the dataset and integrating other features that are excluded from the existing dataset. Additionally, two performance metrics, AUC and ROC, can be added to evaluate the performance of each model. The analysis can also be repeated using the most recent dataset to confirm the usefulness of these predictions in actual business settings. Finally, the model must be installed and integrated into the systems to access real-time data and make predictions in line with it.

REFERENCES

[1] U.S. Department of Transportation, "U.S. Scheduled Service Airline 2021 Passengers Increased 83% from 2020, Decreased 27% from Pre-pandemic 2019 (Preliminary)," Bureau of Transportation Statistics, February, 2022. https://www.bts.gov/newsroom/us-scheduled-service-airline-2021-passengers-increased-83-2020-decreased-27-pre-pandemic.

[2] I. Agarwal and K.R. Gowda, "The effect of airline service quality on customer satisfaction and loyalty in India," Materials Today: Proceedings, vol. 37, pp. 1341–1348, 2021.

[3] T. Noviantoro and J.-P. Huang, "Investigating airline passenger satisfaction: data mining method," Research in Transportation Business & Management, vol. 43, p. 100726, 2022.

[4] S. Tahanisaz and S. Shokuhyar, "Evaluation of passenger satisfaction with service quality: a consecutive method applied to the airline industry," Journal of Air Transport Management, vol. 83, p. 101764, 2020.

[5] C.C.H. Law, Y. Zhang and J. Gow, "Airline service quality, customer satisfaction, and repurchase intention: Laotian air passengers' perspective," Case Studies on Transport Policy, vol. 10, no. 2, pp. 741–750, 2022.

[6] R. Khan and S. Urolagin, "Airline sentiment visualization, consumer loyalty measurement and prediction using Twitter data," International Journal of Advanced Computer Science and Applications, vol. 9, no. 6, pp. 380-388, 2018.

[7] S. Kumar and M. Zymbler, "A machine learning approach to analyze customer satisfaction from airline tweets," Journal of Big Data, vol. 6, no. 1, p. 62, 2019.
[8] F.R. Lucini, L.M. Tonetto, F.S. Fogliatto and M.J. Anzanello, "Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews," Journal of Air Transport Management, vol. 83, p. 101760, 2020.
[9] S. Leon and J.C. Martín, "A fuzzy segmentation analysis of airline passengers in the U.S. based on service satisfaction.," Research in Transportation Business & Management, vol. 37, p. 100550, 2020.
[10] S. Conlon, L. Simmons and F. Liu, "Predicting Tech employee job satisfaction using machine learning techniques Sumali J. Conlon1 Lakisha L. Simmons2 Feng Liu3," International Journal of Management & Information Technology, vol. 16, pp. 97–113, 2021.
[11] L. Yao, Z. Wang, H. Gu, X. Zhao, Y. Chen and L. Liu, "Prediction of Chinese clients' satisfaction with psychotherapy by machine learning," Frontiers in Psychiatry, vol. 14, p. 947081, 2023.
[12] E.M. Polce, K.N. Kunze, M.C. Fu, G.E. Garrigues, B. Forsythe, G.P. Nicholdon, B.J. Cole and N.N. Verma "Development of supervised machine learning algorithms for prediction of satisfaction at 2 years following total shoulder arthroplasty," Journal of Shoulder and Elbow Surgery, vol. 30, no. 6, pp. e290–e299, 2021.
[13] S.F. Sutan Faisal and N. Sutan Faisal, "Implementation of K-nearest neighbor algorithm for customer satisfaction," Buana Information Technology and Computer Sciences, vol. 1, no. 2, pp. 27–32, 2020.
[14] M.G. Bellizzi, L. Eboli, G. Mazzulla and M.N. Postorino, "Classification trees for analysing highly educated people satisfaction with airlines' services," Transport Policy, vol. 116, pp. 199–211, 2022.
[15] Surigao State College of Technology, Surigao City, Philippines and R.E. Talingting, "A data mining-driven model for job satisfaction prediction of School Administrators in DepEd Surigao del Norte Division," International Journal of Advanced Trends in Computer Science and Engineering, vol. 8, no. 3, pp. 556–560, 2019.
[16] S. Lal Mukherjee and S. Dutta, "A comparative study of school parent satisfaction predictors using different classifiers," International Journal for Research in Applied Sciences and Biotechnolgy, vol. 8, no. 3, 2021.
[17] L.H. Poushy, S.A. Bhuiyan, M. Parvin, R.A. Hossain, N.N. Moon, J. Nooder and A. Mahbuba, "Satisfaction prediction of online education in COVID-19 situation using data mining techniques: Bangladesh perspective," IJECE, vol. 12, no. 5, p. 5553, 2022.
[18] S. Aisyah, P. Marpaung, W. Aprinai, K. Saharja, I.M.Y. Suryawan, B.T.A. Nugroho, A. Nurbaedi, H.A. Hasibuan, B. Nadeak and A. Tohir, "Analysis of the effect of the lecturer satisfaction with the Naive Bayes Data Mining technique on institutional performance," Journal of Physics: Conference Series, vol. 1933, no. 1, p. 012034, 2021.
[19] S.S. Roy, D. Kaul, R. Roy, C. Barna, S. Mehta and A. Misra, "Prediction of Customer Satisfaction Using Naive Bayes, MultiClass Classifier, K-Star and IBK," in Soft Computing Applications, vol. 634, V. E. Balas, L. C. Jain, and M. M. Balas, Eds. Cham: Springer International Publishing, 2018, pp. 153–161.
[20] Y. Li, H. Li and H. Yao, "Analysis and study of diabetes follow-up data using a data-mining-based approach in new urban area of Urumqi, Xinjiang, China, 2016-2017," Computational and Mathematical Methods in Medicine, vol. 2018, pp. 1–8, 2018.
[21] M.S. Bouzakraoui, A. Sadiq and A.Y. Alaoui, "Customer satisfaction recognition based on facial expression and machine learning techniques," Advances in Science, Technology and Engineering Systems Journal, vol. 5, no. 4, pp. 594–599, 2020.
[22] V. Jackins, S. Vimal, M. Kaliappan and M.Y. Lee, "AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes," The Journal of Supercomputing, vol.77, pp. 5198-5219, 2020.
[23] L. Zhu, D. Qiu, D. Ergu, C. Ying and K. Liu, "A study on predicting loan default based on the random forest algorithm," Procedia Computer Science, vol. 162, pp. 503–513, 2019.
[24] P. Sharma, S. Banerjee, D. Tiwari and J.C. Patni, "Machine learning model for credit card fraud detection-a comparative analysis," The International Arab Journal of Information Technology, vol. 18, no. 6, pp. 789-796, 2021.
[25] D. Bzdok, M. Krzywinski and N. Altman, "Machine learning: supervised methods," Nature Methods, vol. 15, pp. 5–6, 2018.
[26] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, 1967.
[27] J. Kiani, C. Camp and S. Pezeshk, "On the application of machine learning techniques to derive seismic fragility curves," Computers and Structures, vol. 218, pp. 108–122, 2019.
[28] Y. Chen, E.K. Garcia, M.R. Gupta, A. Rahimi and L. Cazzanti, "Similarity-based classification: concepts and algorithms," Journal of Machine Learning Research, vol. 10, no. 27, pp. 747–776, 2009.
[29] K.Q. Weinberger and L.K. Saul, "Distance metric learning for large margin nearest neighbor classification," Journal of Machine Learning Research, vol. 10, no. 9, pp. 207–244, 2009.
[30] A.A. Srivastava, A. Chaudhary, A. Gautam, D.P. Singh and R. Khan, "Prediction of students performance using knn and decision tree-a machine learning approach," Strad, vol. 7, no. 9, 2020.
[31] A. Gershman, A. Meisels, K.-H. Lüke, L. Rokach, A. Schclar and A. Sturm, "A decision tree based recommender system," in Proceedings of 10th International Conference on Innovative Internet Community Services, Thailand, 2010, pp. 170–179.
[32] S.D. Jadhav and H.P. Channe, "Efficient recommendation system using decision tree classifier and collaborative filtering," International Research Journal of Engineering and Technology, vol. 3, no. 8, pp. 2113–2118, 2016.

[33] B. Gupta, A. Rawat, A. Jain, A. Arora and N. Dhami, "Analysis of various decision tree algorithms for classification in data mining," International Journal of Computer Applications, vol. 163, no. 8, pp. 15–19, 2017.

[34] D. Nibbering and T.J. Hastie, "Multiclass-penalized logistic regression," Computational Statistics & Data Analysis, vol. 169, p. 107414, 2022.

[35] M.R. Özkale and E. Arıcan, "A first-order approximated jackknifed ridge estimator in binary logistic regression," Computational Statistics, vol. 34, pp. 683–712, 2019.

[36] J.I.E. Hoffman, *Biostatistics for Medical and Biomedical Practitioners*, 2nd ed. United State of America: Stacy Masucci, 2019, pp. 581–589.

[37] L. Breiman, "Random Forest," Machine Learning, vol. 45, pp. 5–32, 2001.

[38] A.M. Kalliguddi and A.K. Leboulluec, "Predictive modeling of aircraft flight delay," Universal Journal of Management, vol. 5, no. 10, pp. 485–491, 2017.

[39] L. Breiman, "Bagging Predictors," Machine Learning, vol. 24, pp. 123–140, 1996.

[40] Hartatik, K. Kusrini and A. Budi Prasetio, "Prediction of student graduation with Naive Bayes algorithm," in Proceedings of 2020 Fifth International Conference on Informatics and Computing, Indonesia, 2020, pp. 1–5.

[41] S. Chen, G.I. Webb, L. Liu and X. Ma, "A novel selective naïve Bayes algorithm," Knowledge-Based Systems, vol. 192, p. 105361, 2020.

[42] S. Ranganathan, K. Nakai, C. Shönbach, and M. Gribskov, *Encyclopedia of Bioinformatics and Computational Biology*, United Kingdom: Oliver Walter, 2019, pp. 403–412.

[43] A. Gautam, P. Singh, B. Raman and H. Bhadauria, "Automatic classification of leukocytes using morphological features and Naïve Bayes classifier," in Proceedings of 2016 IEEE Region 10 Conference, Singapore, 2016, pp. 1023–1027.

[44] Y. L, M. Arunadevi and C. P. S. Prakash, "Predicton of MRR & surface roughness in wire EDM machining using decision tree and Naive Bayes algorithm," in Proceedings of 2021 International Conference on Emerging Smart Computing and Informatics, India, 2021, pp. 527–532.

[45] R.E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," Machine Learning, vol. 37, pp. 297–336, 1999.

[46] Y. Freund and R.E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Journal of Computer and System Sciences, vol. 55, pp. 119–139, 1997.

[47] Ankthon, "*Python-seaborn.factorplot() method*," https://www.geeksforgeeks.org/python-seaborn-factorplot-method/

[48] V.R. Joseph, "Optimal ratio for data splitting," Statistical Analysis and Data Mining, vol. 15, no. 4, pp. 531–538, 2022.

[49] scikit-learn, "sklearn.naive_bayes.GaussianNB," https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html

[50] B.H. Hayadi, J.-M. Kim, K. Hulliyah and H.T. Sukmana, "Predicting airline passenger satisfaction with classification algorithms," International Journal of Informatics and Information System, vol. 4, no. 1, pp. 82–94, 2021.

[51] S. Tsafarakis, T. Kokotas and A. Pantouvakis, "A multiple criteria approach for airline passenger satisfaction measurement and service quality improvement," Journal of Air Transport Management, vol. 68, pp. 61–75, 2018.

[52] E. Park, "The role of satisfaction on customer reuse to airline services: an application of big data approaches," Journal of Retailing and Consumer Services, vol. 47, pp. 370–374, 2019.

[53] K. Gao, Y. Yang and X. Qu, "Examining nonlinear and interaction effects of multiple determinants on airline travel satisfaction," Transportation Research Part D: Transport and Environment, vol. 97, p. 102957, 2021.

[54] X. Jiang, Y. Zhang, Y. Li and B. Zhang, "Forecast and analysis of aircraft passenger satisfaction based on RF-RFE-LR model," Scientific Reports, vol. 12, p. 11174, 2022.

[55] M.H. Imam Bijoy, S.A. Akhi, M.A.A. Nayeem, M.M. Rahman and M.J. Mia, "Prediction of internet user satisfaction levels in Bangladesh using data mining and analysis of influential factors," Bulletin EEI, vol. 11, no. 2, pp. 926–935, 2022.

[56] S.F. Shetu, I. Jahan, M.M. Islam, R. Ara Hossain, N.N. Moon and F. Narin Nur, "Predicting satisfaction of online banking system in Bangladesh by machine learning," in Proceedings of 2021 International Conference on Artificial Intelligence and Computer Science Technology, Indonesia, 2021, pp. 223–228.

[57] D. Jain, S. Makkar, L. Jindal and M. Gupta, "Uncovering employee job satisfaction using machine learning: a case study of Om Logistics Ltd," in Proceedings of International Conference on Innovative Computing and Communications, Singapore, vol. 2, 2021, pp. 365–376.