Data Mining for Airline Industry: Investigating satisfaction of airline passengers

MSc Research Project
MSc. Data Analytics

Tejas Mahesh Jadhav
Student ID: x20206861

School of Computing
National College of Ireland

Supervisor: Mr. Abdul Razzaq

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Tejas Mahesh Jadhav
Student ID:	X20206861
Programme:	MSc. Data Analytics
Year:	2022-2023
Module:	MSc. Research Project
Supervisor	Mr. Abdul Razzaq
Submission Due Date:	01/02/2023
Project Title:	Data Mining for Airline Industry: Investigating satisfaction of airline passengers
Word Count:	5902
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Data Mining for Airline Industry: Investigating satisfaction of airline passengers

Tejas Mahesh Jadhav
x20206861

Abstract

In the robust competition of the aviation industry, passenger satisfaction plays a crucial role in the growth and success of an airline company. Air carriers that are able to identify and satisfy passenger demand stand out in the market with increased sales and more loyal customers. Passengers these days not only consider the ticket prices, but also scrutinize the quality of the airline services before selecting their carrier. Hence it is essential for the airline companies to provide quality services to their passengers. The feedback and reviews given by the passengers can act as a useful tool in understanding their service expectations and demands of the passengers. With the help of data mining techniques, airline companies can not only gauge the satisfaction of their passengers but also discover insights for improvising the services. Apart from conventional feedback forms, social media platforms like Twitter are becoming a popular choice among passengers to express their views and feedback on their travels. These feedbacks can also be used by the airline companies for understanding the sentiments of the passengers. To provide a direction on the same front, this study investigates the satisfaction of airline passengers by leveraging data mining techniques on two different datasets. The first dataset is a survey dataset consisting of satisfaction scores of 103903 airline passengers over different services of airline companies. The second dataset consists of 14641 tweets of the airline passengers. After evaluating 18 different classifiers on both the datasets, it was observed that Soft Voting Classifier is a better performing classifier for the first dataset and gives an accuracy of 96.60 percent. For the second dataset, the stacking classifier is a better performing classifier which showcased an accuracy of 93.98 percent. As for service attributes, in-flight Wi-Fi services and online boarding provision have a greater impact on passenger’s satisfaction, hence airline companies should focus on improving those services aspects. In addition, the ‘Type of Travel’ is also an important aspect of passenger’s satisfaction. Hence, the airline crew should behave accordingly with the passengers traveling for business or personal travel purposes.

1 Introduction

The competition in the Aviation Industry has become very robust these days (Noviantoro and Huang; 2022). Airline companies across the world have created extraordinary levels of affordability and accessibility at all levels (Noviantoro and Huang; 2022; Wan and Gao; 2015). Air travel, which once used to be an exclusive opportunity, has now become a first choice for the majority (Noviantoro and Huang; 2022). Over the years, new carriers started entering the aviation market and the existing started expanding their footprints (Noviantoro and Huang; 2022; Wan and Gao; 2015; Aljedaani et al.; 2022). In the aviation business, passenger's satisfaction and loyalty plays a crucial role in the growth and success of an airline company. Air carriers that are able to identify and satisfy passenger's demand stand out in the market with increased sales and loyal customers.

Reduction of ticket prices is one of the key strategies of airline carriers to attract passengers (Noviantoro and Huang; 2022). However, this price reduction strategy is not efficient in the longer run (Chang and Yeh; 2002). According to (Punel et al.; 2019), passengers these days not only consider the ticket price but also scrutinize and compare the quality of the services from airline to airline when selecting their carrier. Several studies also suggested that quality of the airline services has a significant impact on passenger's satisfaction (Punel et al.; 2019) (Sezgen et al.; 2019; RaE et al.; 2022), passengers' loyalty ` (Chonsalasin et al.; 2022; RaE et al.; 2022; Singh; 2021), and choice of airline carrier (Jo- ` har et al.; 2020; Adeola and Adebiyi; 2014). Hence it is essential for the airline companies to provide quality services to their passengers.

The feedback and reviews given by the passengers can act as a useful tool in understanding the service expectations and demands of the passengers (Noviantoro and Huang; 2022; Wan and Gao; 2015; Aljedaani et al.; 2022). By adopting data mining and machine learning techniques, aviation companies can not only be able to gauge the satisfaction levels of their passengers but also come across insights that can be used for improvising the passenger's satisfaction. Apart from conventional feedback forms, social media platforms like Twitter and Facebook are becoming popular among passengers to express their views and feedback on their travels (Wan and Gao; 2015). The feedback from these social media platforms can also be leveraged by the airline companies for understanding the sentiments of their passengers (Wan and Gao; 2015; Aljedaani et al.; 2022).

To provide a direction on the same front and assist airline companies in investigating the satisfaction of customers, this study examines the satisfaction of airline passengers by leveraging data mining techniques on two different datasets. The first dataset is a survey dataset consisting of satisfaction scores of 103903 airline passengers over different services of airline companies. The second dataset consists of 14641 tweets of the airline passengers. To obtain better prediction results, this study implemented several machine learning, deep learning and ensemble techniques. Some of the classification algorithms employed for this study include Random Forest, Decision Tree, Gradient Boosting, KNN, Neural Network, Multi-Layer perceptron, Logistic Regression, Ensemble Stacking and Voting etc. Overall 18 different classifications models were developed for each of the datasets.

Prior studies have utilized the same datasets (First dataset - (Noviantoro and Huang; 2022; Tan; 2021; Gao et al.; 2021) - Second dataset -(Gupta et al.; 2021; Tusar and Islam; 2021; Rane and Kumar; 2018; Wan and Gao; 2015; Aljedaani et al.; 2022; Kalangi et al.; 2021; Qudar and Mago; 2020; Rustam et al.; 2019; Kamruzzaman et al.; 2021)) with different machine learning and deep learning techniques. In contrast to the previous studies, this study explores the performance of stacking and voting ensemble techniques in classifying the sentiments of the airline passengers. In addition, this study also compares the performance of the classifiers on two different datasets. To the best of knowledge, no previous studies have implemented ensemble stacking with base classifiers as - Random Forest, XGBoost, Extra Trees Classifier, LightGBM and Meta classifier as - Logistic Regression and Multilayer Perceptron and voting classifier with Random Forest, Gradient Boosting, XGBoost, Extra Trees Classifier, Support Vector Machine, Multi-Layer Perceptron, LightGBM and Logistic Regression. In addition, to the best of knowledge none of the studies provided an empirical comparison of different classifiers on the mentioned datasets.

Specifically, the following research question is addressed in this study

”Do the stacking and voting techniques perform better in classifying the satisfaction of airline passengers?”

To address the above research question, following schema were formulated:

(i). Adopting a data mining methodology for deriving insights from the data and building various classification models.
(ii). Evaluating and comparing the performance of the models.
(iii). Comparing the performance of the models with several recent approaches on classification of airline passenger’s satisfaction.

The main contributions of this work include:

- An empirical comparison of different machine learning, neural network and ensemble classifiers on two different datasets.
- Deriving the important attributes that needs to be focused upon to increase the passenger satisfaction.
- Two different stacking and voting classifiers for classifying the satisfaction of airline passengers

Rest of the study is compiled into following four sections:
Section 2: Presents the related studies related to the problem statement
Section 3: Presents methodology and model specifications of this study
Section 4: Evaluates the performance of the classifiers
Section 5: Discusses the results and concludes the study with directions for future work.

2 Related Work

Various studies have explored different data mining techniques and machine learning approaches for classifying the satisfaction of airline passengers. The related studies can be classified into following categories: Importance of service quality in airline industry, Machine Learning and Deep Learning techniques for predicting the satisfaction airline passengers and Utilization of Twitter data for gauging the satisfaction of airline passengers. In addition, this section also discusses the application of stacking and voting techniques for solving classification problems in the last two sub-sections.

2.1 Importance of service quality in airline industry

The quality of the airline services have a significant influence on passenger's satisfaction and loyalty behavior (Sezgen et al.; 2019; Gupta et al.; 2021; RaE et al.; 2022). As a part ` of an empirical study, (RaE et al.; 2022) presented four important variables which can be ` used for modeling and analyzing the quality of airline services. These four variables are: Pre-flight ground service quality (PFGSQ), In-flight service quality (IFSQ), Post-flight ground service quality (PFSQ) and Overall Service Quality (OSQ) (RaE et al.; 2022). ` Out of these four variables, the first variable encapsulates ground services such as airport lounges, baggage service, provisions of check-in services and on-ground staff's behavior. The second variable IFSQ takes into account the flight's catering services, passenger assistance, seat comfort and the conduct of the flight crew. Similar to the first variable, PFSQ summarizes the on ground services post the flight is landed. It helps in scrutinizing the disembarkation procedure, baggage handling and return services, facilities of waiting areas, transportation and hotel services. The last variable OSQ, takes into account the quality of the services over the whole process. From the results of the reliability analysis (RaE et al.; 2022), it was observed that all these variables can be useful in scrutinizing ` and perceiving the quality of the airline services. In addition (RaE et al.; 2022) also ` pointed out that the satisfaction and loyalty of the passenger is significantly impacted by the quality of the services. Better the quality of the services, better is the customer satisfaction index and customer retention rate.

Drilling further down on the subject of airline passenger satisfaction, (Sezgen et al.; 2019) investigated the key factors of satisfaction for different service classes and business models of the airline industry. After implementing the Latent Semantic Analysis (LSA) method on over five thousands online reviews of the airline passengers, (Sezgen et al.; 2019)] came to a conclusion that - the code of conduct and friendliness of the flight crew is an essential factor for the passengers traveling in the Economy class, whereas for the Premium cabin passengers the overall experience of the fight journey is a deciding factor. For passengers who chose to fly via low cost carriers, the price of the flight ticket is an important factor of satisfaction for them. In addition to these insights, (Sezgen et al.; 2019) also uncovered that services such as luggage disruption, flight delay and cancellation, seat comfort and legroom and staff behavior are major reasons for passengers dissatisfaction among all service classes.

In this multicultural world, opinions and perceptions of the individuals over a certain aspect vary from place to place due to different geographical and cultural influences (Punel et al.; 2019). Similar traits can be found in the airline industry. To understand and compare the variation in the opinions of the passengers across different geographical regions, (Punel et al.; 2019) developed a framework called Path Analysis and carried out the analysis on over 40,000 reviews of the airline passengers. The results of the analysis confirm the hypothesis that geographical regions, country of residence have a great influence on passenger's perceptions, experience and evaluation of the airline services. In addition to this, (Punel et al.;

2019) also came across some interesting insights like - the North American passengers are more concerned about the cost of air traveling rather than the in-flight services (Punel et al.; 2019). The passengers from the Asian countries prefer Asian airlines over North American airlines. Seat comfort along with the staff service is an important factor for all the passengers across the globe. The results of the analysis also confirms that the expectations of the passengers vary from class to class. The first and business class passengers are more concerned about food and beverages and in-flight entertainment, whereas the economy class passengers are more anxious regarding the value for Money (Punel et al.; 2019).

The aviation industry took a major hit in the year 2020 and 2021 due to the outbreak of COVID-19 pandemic (Zhang et al.; 2021). The number of flights and passengers had dropped significantly during these two years. Even when the travel restrictions were uplifted for a brief period of time, the industry did not see much of an improvement. In addition to this, the fear of the pandemic and the control measures had caused remarkable changes in the travel behavior of the airline passengers. To further understand the pandemic-induced psychological changes, (Zhang et al.; 2021) carried out an analysis on the passenger data and found that because of the restrictions and control measures of the pandemic, passengers are now arriving 100-180 minutes earlier a t the airports than usual. The refund rates spiked during the year 2020 and 2021. In addition to this, the aviation industry saw a great decline in the numbers of children, middle-aged and elderly passengers (Zhang et al.; 2021). In order to recover in the post-pandemic era, it is essential for the airline industry to upgrade and formulate services to the passengers taking into consideration the pandemic-induced psychological and behavioral changes of the passengers (Zhang et al.; 2021).

In any business sector, customer’s loyalty plays a significant role in the growth and success of a company (Singh; 2021). Loyalty causes more repeat purchases, more positive word-of-mouth in the market and thus more market share and better financial performance (Singh; 2021). Thus it is essential for the companies to identify their loyal customers and look after them. Just like any other business sector, companies in the aviation industry strive hard towards maintaining and increasing the count of loyal customers in the market (Singh; 2021). To provide a direction on the same front, (Chonsalasin et al.; 2022) developed a structural equation model for identifying the loyal passengers of the airline industry. The results of the developed model suggested that the quality of the airline services had a significant influence in the loyalty of the passengers. Another study (Chonsalasin et al.; 2022) employed Artificial Neural Network for classifying loyal airline customers. The developed model showcased an accuracy of 89 percent and was efficiently able to address the nonlinear relationships of the predicting factors, but it lacked in acknowledging the causal relationship among the variables. So to address the causal relationship, (Chonsalasin et al.; 2022) developed a structural equation model. The findings of the structural model suggested that - passenger satisfaction, quality of the airline service and passenger perceived value are essential factors contributing towards passengers loyalty (Chonsalasin et al.; 2022).

2.2 Importance of service quality in airline industry

Service Quality is one of the most essential factors for attracting passengers (Noviantoro and Huang; 2022; Sezgen et al.; 2019; Gupta et al.; 2021). Hence it is important for the airline companies to keep a track of their services and improve the essential attributes (Noviantoro and Huang; 2022; RaE et al.; 2022; Sezgen et al.; 2019). To assist the airlines ` companies in assessing the impact of their services on passenger's satisfaction, (Noviantoro and Huang; 2022) exploited the data mining techniques on a airline passenger dataset and came to a conclusion that baggage handling, online boarding, inflight Wi-Fi and entertainment are the critical aspects for passengers satisfaction and the airline companies should give more attention on improving these four services. In addition, (Noviantoro and Huang; 2022) after experimenting with machine learning and deep learning techniques like Random Forest, Decision Tree, Naive Bayes, KNN, SVM and Neural Network came to a conclusion that Neural Networks are the one of the most fittest techniques when it comes to the prediction of airline passengers satisfaction (Noviantoro and Huang; 2022). Another study (Tan; 2021), after experimenting with multiple machine learning and ensemble techniques suggested that Random Forest is the best technique for predicting the satisfaction of the airline passenger. (Tan; 2021) also suggested that airlines should focus on services like seat comfort, online boarding, inflight Wi-Fi and entertainment services for improving the satisfaction of their passengers.

For assisting the airline companies in designing cost-effective and efficient services measures, (Gao et al.; 2021) examined the complex interaction between the passenger satisfaction and services attributes and found that passenger attributes such as customer type, type of travel and traveling class along with airline service such as online boarding, inflight Wi-Fi and baggage handling are critical factors that have larger impact passengers satisfaction. (Gao et al.; 2021) also came across some interesting insights like - providing a low-quality Wi-Fi service may not have much impact on the passengers satisfaction compare to having no Wi-Fi service at all, also a good online boarding service is more important to passengers of eco and eco plus class compared to business class (Gao et al.; 2021). In addition, (Gao et al.; 2021) also suggested that machine learning algorithms such SVM, Random Forest and MLP showcase better performance in predicting the satisfaction of the airline passengers compared to the conventional logistic regression model (Gao et al.; 2021).

Customer's feedback and reviews play an important role in assessing the performance of a company's product and services (Sharma et al.; 2021) and can also be used for improving and reshaping the products and existing services (Hasan et al.; 2020). With the help of data mining techniques, customer's feedback and reviews can be processed and the polarity of their feedback can be categorized as satisfaction or dissatisfaction, or positive or negative (Hasan et al.; 2020). To stay competitive and create a name in the aviation industry, it is very essential for an airline company to take into consideration the feedback of the passenger and improve their services as per their demand (Noviantoro and Huang; 2022).

2.3 Importance of service quality in airline industry

Nowadays, many customers across the globe prefer to review the critics and opinion of the people on the online platforms before trying new products and services. Twitter and Facebook are one such platforms where people post their reviews and feedback. All those online reviews of the consumers can be harnessed by the company to evaluate the performance of their products in the market and discover areas of improvement. Sentiment

Analysis with machine learning and deep learning is one such approach which can be used by the companies for assessing the online reviews of the consumers. To provide a direction on the same front and assist the airline companies in evaluating the online reviews of their passengers, (Tusar and Islam; 2021) conducted a sentiment analysis on 14640 tweets of the airline passengers and compared the performance - SVM, Multinomial Naive Bayes, RF and LR algorithm for classifying the sentiments of the tweets. The results of the experiment showcased that SVM and LR are better performing candidates in classifying the tweets of the airline passenger. Along with that (Tusar and Islam; 2021) also pointed out these two algorithms perform slightly better when texts are processed with BoW technique.

Deep learning is another popular class of prediction technique which is widely used for understanding the sentiments behind a text (Gupta et al.; 2021). For evaluating how well the deep learning techniques can perform well in classifying the sentiments of a tweet, (Gupta et al.; 2021) developed a neural network model for classifying the sentiments of airline related tweets. After evaluating the performance of the neural network against machine learning technique, (Gupta et al.; 2021) concluded that a neural network model along with dropout layers is a better choice for classifying the sentiments of the airline passengers compared to the machine learning algorithms (Gupta et al.; 2021).

Along with the machine learning algorithms, (Rane and Kumar; 2018) explored the application of boosting algorithms for understanding the tweet sentiments of the passen ger. The evaluation results of (Rane and Kumar; 2018) suggested that the AdaBoost algorithm is a better performing candidate compared to the machine learning algorithms like DT, RF, SVM, Gaussian Naive Bayes, LR and KNN. Contrast to (Rane and Kumar; 2018), the study (Kalangi et al.; 2021) suggests that Naive Bayes algorithm gives more accurate results in classifying the sentiment of airline passengers.

Another paper (Aljedaani et al.; 2022) experimented with deep learning and machine learning techniques for classifying the tweets of the airline passengers. The LSTM and LSTM-GRU developed by (Aljedaani et al.; 2022) showcased the highest accuracy of 0.97 in classifying the tweets of the airline passengers. In addition, the experiments conducted by (Aljedaani et al.; 2022) also suggested that the performance of the machine learning techniques like SVM and ETC becomes better when text processing is conducted using TF-IDF and BoW techniques.

Nowadays pre-trained models are becoming a popular choice in the domain of sentiment analysis (Qudar and Mago; 2020). Pre-trained models are trained on millions of data points and can be used for classifying new content right away with slight configurations (Qudar and Mago; 2020). To explore the performance of the pre-trained model in classifying the sentiments of the airline passengers, (Qudar and Mago; 2020) created two versions of the pre-trained model - TweetBERT. The models proposed by (Gupta et al.; 2021) outperformed the state-of-the-art models like BERT, Biobert, SciBERT, RoBERTa and Albert in classifying the sentiments of the passengers with an accuracy of 92.99 percent (Qudar and Mago; 2020).

2.4 Ensemble Stacking technique for making predictions

Considering the inabilities of a single model (Li et al.; 2019), many studies and researchers resort to stacking techniques for solving a problem (Li et al.; 2019). Stacking technique combines multiple models and harnesses their capabilities to give better results than a single model. Taking into account the benefits of stacking, (Li et al.; 2019) developed a stacking model for detecting phishing webpages using the GBDT, LightGBM and XG Boost models (Li et al.; 2019). The proposed stacked model outperformed the individual machine learning models with an accuracy of 97.30 percent and 98.60 percent on the 50K-PD and 50K-IPD dataset respectively.

Another study (Neshir et al.; 2021) owing to the benefits of stacking strategy developed a stacked model using the RF,SVM,NB and LR for classifying the reviews of Amharic speaking customers on online products and services. The stacked model proposed by (Neshir et al.; 2021) achieved an accuracy of 90 percent surpassing the performance scores of other individual machine learning models.

Another study (Pavlyshenko; 2018) explored the performance of stacking techniques on two different datasets. For the dataset of sales-time series forecasting, (Pavlyshenko; 2018) developed a stacked model using Extra Tree, Linear Regression, NN and XGBoost and for the dataset of Classification of detective parts, (Pavlyshenko; 2018) developed a stacked model using the Logistic regression, Naive Bayes and Xgboost. The experiment results of both the case-studies concluded that the stacking technique is a better option when the objective is to improve performance of the prediction models (Pavlyshenko; 2018).

Another study (Rustam et al.; 2019), proposed a novel approach based on the stacking technique to classify the sentiments of the Twitter Posts. To enhance the sentiment analysis (Rustam et al.; 2019) integrated a lexical dictionary with a stacked model. The stacked model developed consisted of three LSTM models as base classifiers and Logistic Regression as a Meta classifier (Rustam et al.; 2019). The proposed approach outperformed the conventional machine learning models like logistic regression, AdaBoost, and random forest and with an accuracy score of 99 percent.

2.5 Voting Classifiers for making predictions

Voting Classifier is a powerful technique which has the ability to overcome the biases of the models and derive a generalized fit (Rustam et al.; 2019; Kamruzzaman et al.; 2021). By integrating the benefits of logistic regression and naive bayes, (Kamruzzaman et al.; 2021) developed a hard voting classifier for classifying the sentiments of the airline passengers. However, the voting classifier developed by (Kamruzzaman et al.; 2021) could not outperform the performance of the individual Logistic and Niave Bayes classifier. Another study (Punel et al.; 2019) attempted to classify the sentiments of the airline passengers using the soft voting classifier model. The voting classifier developed by (Rustam et al.; 2019), combined the classification probabilities of logistic regression and stochastic gradient descent model to classify the sentiments. The soft voting model achieved an accuracy of 79.1 percent and surpassed the performance of the individual logistic and stochastic gradient descent model. In addition to classifying the sentiments of the customers, various studies have achieved better results with voting models in other fields. (Khuriwal and Mishra; 2018) proposed a soft voting classifier which combined the results from logistic regression and neural network model for detecting breast cancer. The voting model developed by (Khuriwal and Mishra; 2018) achieved an accuracy of 98 percent, and surpassed the performance of the individual logistic and neural network model (Khuriwal and Mishra; 2018). Another study (Atallah and Al-Mousa; 2019) developed a hard voting classifier for solving the problem of heart disease detection (Atallah and Al-Mousa; 2019). The voting classifier developed by (Atallah and Al-Mousa; 2019), combined the votes from KNN, Random Forest, Logistic Regression and Stochastic Gradient Descent model and gave an accuracy of 90 percent, exceeding the accuracy of the individual classifier (Atallah and Al-Mousa; 2019).

Various studies have utilized different machine learning and neural network techniques for classifying the satisfaction of airline passengers. However, none of the studies have provided a rigorous comparison of different machine learning, neural network and ensemble classifiers on the mentioned datasets. This study overcomes this limitation.

3 Research Methodology

It is essential for a data mining project to follow a certain methodology (Schroer et al.; 2021; Moro et al.; 2011). Methodologies act as guidelines, and make sure that none of the crucial steps are skipped during the development of a project (Ayele; 2020; Schroer et al.; 2021). To answer the research question and achieve the project objectives, a custom methodology inspired from the CRISP-DM methodology was adopted for this study. Figure 1. showcases the steps and flow of the research methodology.

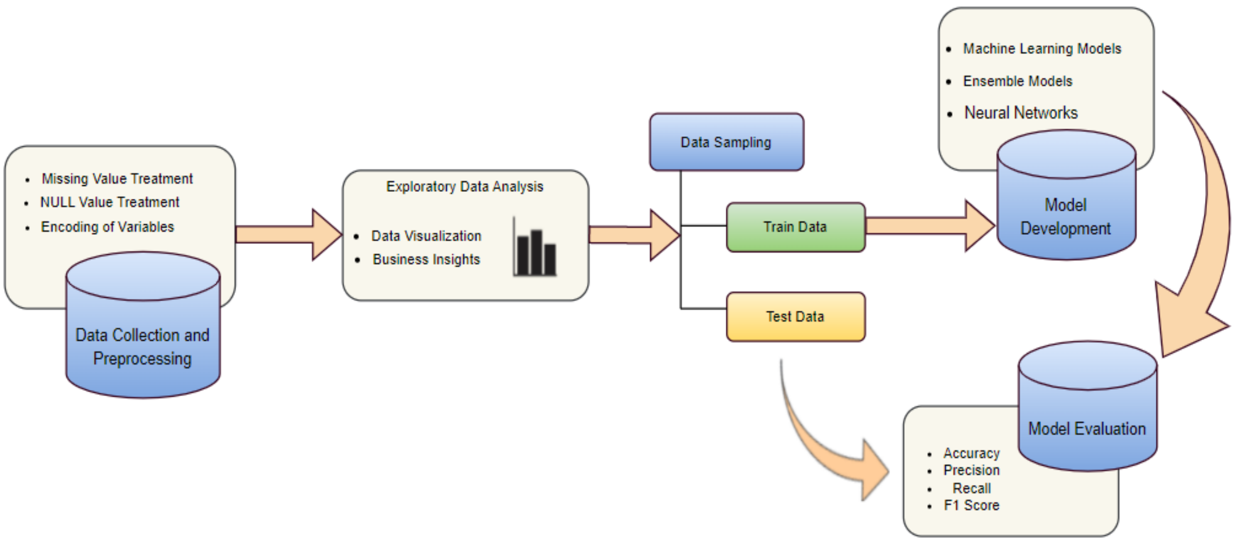

Figure 1: Methodology

3.1 Data Collection and Pre-processing

In this study, two data were utilized. The first dataset[1] is a survey dataset consisting of satisfaction scores of 103903 airline passengers over different services of airline companies. The second dataset[2] consists of 14641 tweets of the airline passengers. Both the datasets are

[1] https://www.kaggle.com/datasets/deltasierra452/airline-pax-satisfaction-survey
[2] https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment

open source and were obtained from the Kaggle Data Repository. Figure 2. and Figure 3. describe the features present in the dataset.

Attribute	Description
Id	Identification number of the passenger
Gender	Male or Female
Customer Type	Loyal or Disloyal
Age	Age of the passenger
Type of Travel	Business Travel or Personal Personal
Class	Business, Eco Plus or Eco
Flight Distance	Journey Distance
Inflight_wifi Service	Satisfaction Score between 0-5
Departure/Arrival Time	Satisfaction Score between 0-5
Ease of Online Booking	Satisfaction Score between 0-5
Gate Location	Satisfaction Score between 0-5
Food and Drinks	Satisfaction Score between 0-5
Online Boarding	Satisfaction Score between 0-5
Seat Comfort	Satisfaction Score between 0-5
Inflight Entertainment	Satisfaction Score between 0-5
On-board Service	Satisfaction Score between 0-5
Leg room	Satisfaction Score between 0-5
Baggage handling	Satisfaction Score between 0-5
Checkin Service	Satisfaction Score between 0-5
Inflight Service	Satisfaction Score between 0-5
Cleanliness	Satisfaction Score between 0-5
Departure Delay	Delay in departure (in min.)
Arrival Delay	Delay in arrival (in min.)
Satisfaction (Target Varaible)	Satisfied' or 'Neutral or Dissatisfied'

Figure 2: Features present in the first dataset

Attribute	Description
tweet_id	Identification number of the tweet
airline_sentiment (Target Variable)	Positive, Negative, Neutral
airline_sentiment_confidence	Between 0 and 1
negativereason	Reason for the negative sentiment
negativereason_confidence	Between 0 and 1
airline	Name of the Airline
airline_sentiment_gold	NULL for all the rows
name	tweeter_name
negativereason_gold	NULL for all the rows
Gate Location	Satisfaction Score between 0-5
retweet_count	Count for the retweet
text	Actual tweet content
tweet_coord	Location Coordinates where the tweet was generated
tweet_created	Time Stamp of the created tweet
tweet_location	City where the tweet was generated
user_timezone	Time Zone

Figure 3: Features present in the second dataset

As a part of data preprocessing, numerous data operations were carried out on both the datasets. The cleaned versions of both the datasets were then utilized for exploratory data analysis and for developing the classification models. Figure 4. and Figure 5. details the data operations that were conducted on the first and second dataset respectively.

Sr. No.	Operations Carried out on Dataset 1
1.	Treatment of the Missing Values present in the column 'Arrival in delay' (missing values were replaced with median)
2.	Column - ID was dropped
3.	Categorical variables were Encoded using LabelEncoder
4.	All the numerical values were standardized to standard scale

Figure 4: Preprocessing done on the First Dataset

Sr. No.	Operations Carried out on Dataset 2
1.	All the rows with 'neutral' sentiment were removed
2.	Removed - urls, tags, stopwards, punctuations from 'text' column data
3.	Converted 'text' data to numerical metrics using TFIDF
4	Encoded the target variable
5.	Only the 'text' and ailrine_sentiment' columns were used for classification

Figure 5: Preprocessing done on the Second Dataset

As mentioned in Figure 5, all the records with airline sentiment as - neutral were removed from the second dataset and the classification models that were developed for this dataset were binary classifiers capable of classifying the sentiments as - positive or negative.

3.2 Exploratory Data Analysis

Exploratory Data Analysis helps in uncovering business insights and hidden pattern from a dataset. During the exploratory data analysis, the data is visualized using different charts and plots.

After conducting Exploratory Data Analysis on the first dataset, following insights were derived:

(i). About 56 percent of the total passengers either felt dissatisfied or had neural opinion about the services of the airline company.

(ii). Higher dissatisfaction was observed among the passengers of the Economy class.

(iii). People traveling for business purposes felt much more satisfied compared to the passengers traveling for personal purposes.

(iv). A higher rate of dissatisfaction can also be observed in case of the disloyal passengers.
(v). An equal distribution in the satisfaction rate was observed for both men and women.
(vi). Passengers with higher satisfaction fall in the age group of 35 to 60, whereas passengers with unsatisfied feelings fall in the group of 20 to 40.

The EDA chart for the first dataset can be seen in the Figure 6. In addition the table in the Figure 7. showcases the average satisfaction scores of the passengers for different service attributes.

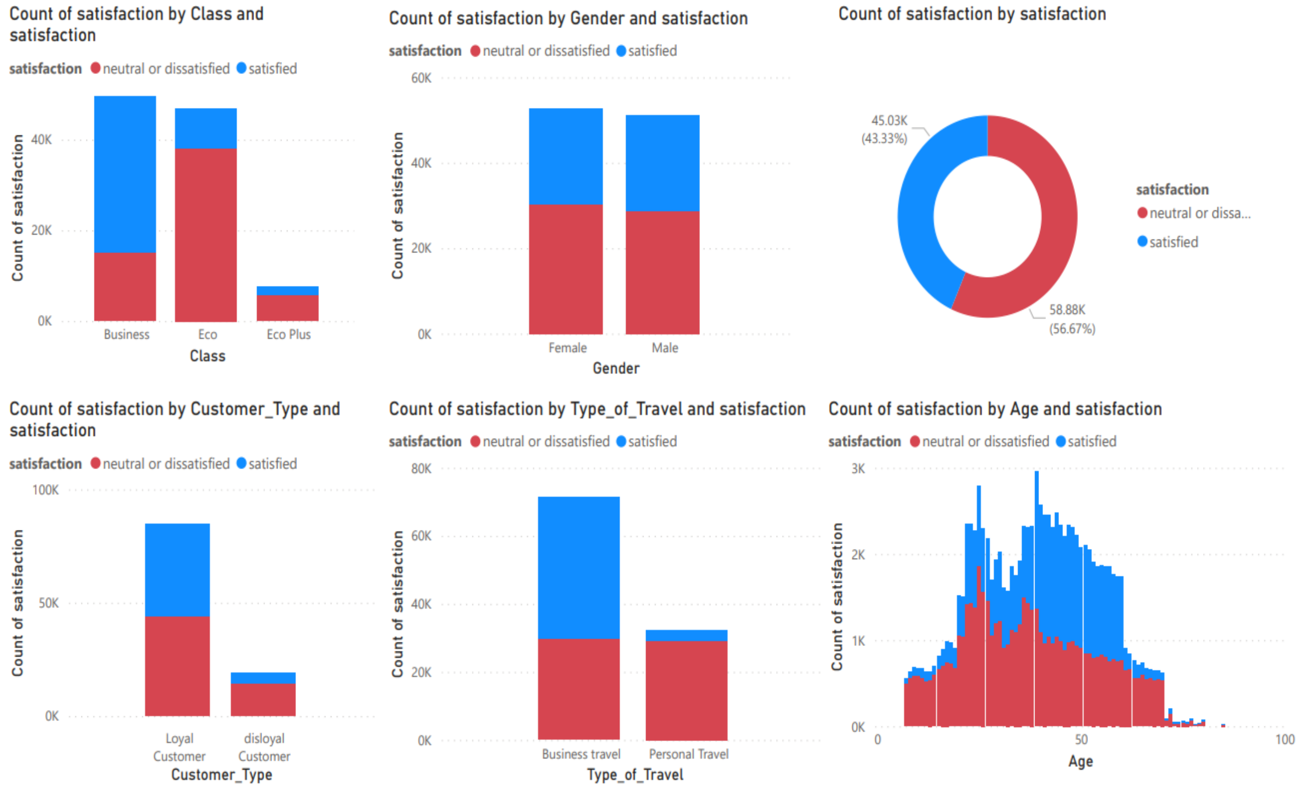

Figure 6: EDA for First dataset

Sr. No.	Service Attributes	Average Score
1	Baggage Handling	3.63
2	Cleanliness	3.29
3	Food and Drinks	3.2
4	Inflight Service	3.6
5	Checkin Service	3.3
6	Gate Location	2.98
7	Inflight Wifi Service	2.73
8	Online Booking	2.76
9	Inflight Entertainment	3.36
10	Leg Room	3.35
11	On-board Service	3.38
12	Online Boarding	3.25
13	Seat Comfort	3.44

Figure 7: Average Satisfaction scores over different service aspects

For the second dataset, word clouds were created for the positive and negative sentiment. Figure 8 showcases the word clouds. From the word clouds, it can be seen that passengers with negative sentiment were unsatisfied with following services of the airline companies: customer service, baggage gate location, seat comfort delayed flight ticket price etc.

(i) Positve Sentiments

(ii) Negative Sentiments

Figure 8: Word Cloud for the second dataset

3.3 Data Sampling

During this step, the dataset is divided into two parts. Out of which, the first part is utilized for building and training the classification model and the second part is then used for testing and evaluating the performance of the model.

For this study, the first dataset was split in the ratio of 80:20. 80 percent of the first data was then used for developing classification models, while the remaining 20 percent was then used for accessing the performance of the models.

During the exploratory data analysis stage, it was observed that the target variable of the second dataset was unevenly distributed. If a classification model is developed on an imbalanced dataset, it will tend to have a bias towards the majority class (Moro et al.; 2011; Fernandez et al.; 2018). Thus to avoid consequences of an imbalanced dataset, SMOTE technique was utilized on the second data. SMOTE balances the distribution of the target classes by increasing the count of minority classes by replicating some of its examples (Moro et al.; 2011; Fernandez et al.; 2018). The distribution of target classes, before and after the SMOTE technique can be seen in the Figure 9.

Before SMOTE	After SMOTE
Negative - 9178 , Positive - 2363	Negative - 9178,Positive - 9178

Figure 9: Class Distribution Before and After SMOTE

3.4 Model Development

Model development is one of the critical stages of a data mining project. During this stage, different prediction and classification models are developed for solving the concerned business problem. To meet the project objectives, a total of 18 different classification models were developed for each dataset. Figure 10 details the specification of the models developed for dataset 1 and 2.

Sr. No.	Technique	Parameters for Dataset 1	Parameters for Dataset 2
1	Random Forest	max_depth=24, random_state=42	max_depth=80, random_state=42
2	Decision Tree	random_state=42	random_state=42
3	Gradient Boosting Classifier	estimators=100,max_depth=8, random_state=42	estimators=100,max_depth=20, random_state=42
4	KNN	n_neighbors=15	n_neighbors=2
5	Naive Bayes	default parameters	default parameters
6	Logistic Regression	penalty='l2',random_state=42	penalty='l2',random_state=42
7	SVM	kernel = 'rbf',gamma='auto',probability=True	probability=True
8	AdaBoost	n_estimators=125, random_state=42	n_estimators=500, random_state=42
9	XGBoost	n_estimators=1000,max_depth=5	n_estimators=1000,max_depth=20
10	LightGBM	objective="binary", n_estimators=100	objective="binary", n_estimators=1000
11	Catboost	iterations=50, depth=8, learning_rate=0.1, loss_function='RMSE'	iterations=50, depth=10, learning_rate=0.1, loss_function='RMSE'
12	Extra Tress Classifier	n_estimators=100, random_state=42	n_estimators=500, random_state=42
13	Neural Network	Layer 1 - dense (Dense),Shape - (None, 24) Layer 2 - dense_1 (Dense),Shape- (None, 11) Layer 3 -dense_2 (Dense),Shape-(None, 1)	Layer 1 - embedding_4 (Embedding),Shape - (None, 21, 128) Layer 2 - dense_12 (Dense),Shape - (None, 21, 128) Layer 3 - dropout_8 (Dropout),Shape - (None, 21, 128) Layer 4 - dense_13 (Dense),Shape - (None, 21, 128) Layer 5 - dropout_9 (Dropout),Shape - (None, 21, 128) Layer 6 - flatten_4 (Flatten),Shape - (None, 2688) Layer 7 - dense_14 (Dense),Shape - (None, 2)
14	MultiLayer Perceptron	hidden_layer_sizes=(24,14),activation = 'relu', random_state=5, verbose=True,learning_rate_init=0.01	hidden_layer_sizes=(15,12),activation = 'relu', random_state=5, verbose=True,learning_rate_init=0.01
15	Stacking_ml	Base Classifiers - Random Forest, XgBoost,Extra Tress Classifier, Meta Classifier - Logistic Regression	Base Classifiers - LightGBM, XgBoost,Extra Trees Classifier, Meta Classifier - Logistic Regression
16	Stacking_ml_mlp	Base Classifiers - Random Forest,XgBoost,Extra Tress Classifier, Meta Classifier - MLP	Base Classifiers - LightGBM,XgBoost,Extra Trees Classifier, Meta Classifier - MLP
17	Hard Voting	Classifiers - Random Forest, Gradient Boosting,XgBoost, Extra Trees Classifier,SVM,MLP Voting - 'hard'	Classifiers - Random Forest, Gardient Boosting,XgBoost, Extra Tree Classifier, SVM,MLP,Logistic,LightGBM Voting - 'hard'
18	Soft Voting	Classifiers - Random Forest, Gradient Boosting,XgBoost, Extra Trees Classifier,SVM,MLP Voting - 'soft'	Classifiers - Random Forest, Gardient Boosting,XgBoost, Extra Tree Classifier, SVM,MLP,Logistic,LightGBM Voting - 'soft'

Figure 10: Model Parameter

3.5 Model Evaluation

Model Evaluation is a mandatory step for a data mining project (Moro et al.; 2011; Fernandez et al.; 2018). During this stage, the efficacy of the models is scrutinized using standard metrics like accuracy, precision, f1 measure etc. Scores from such evaluation metrics are useful in assessing the performance of the models and also for determining the best model that represents our data. For this study, following metrics were employed for gauging the performance of the developed models: Accuracy, Precision, Recall and F1 Score

4 Evaluation

Performance Evaluation is an essential part of the data mining project. It assists in selecting the right model and fitting model for the problem statement and concerned dataset (Moro et al.; 2011; Fernandez et al.; 2018). In this study, the performance of the model was evaluated in four stages as follows:

4.1 Evaluation Scores

Post the development of the models, evaluation scores for all the models were derived. Figure 11. and Figure 12. showcases the evaluation results of the models that were developed on the first and second dataset respectively.

Sr. No.	Technique	Accuracy	Precision	Recall	F1 Score
1	Random Forest	0.9631	0.9735	0.9395	0.9562
2	Decision Tree	0.9461	0.9336	0.941	0.9373
3	Gradient Boosting Classifier	0.9637	0.9732	0.9412	0.9569
4	KNN	0.9271	0.9561	0.8698	0.9109
5	Naive Bayes	0.8642	0.8579	0.8182	0.8376
6	Logistic Regression	0.8725	0.8635	0.8339	0.8484
7	SVMC	0.9552	0.9605	0.9337	0.9469
8	Adaboost	0.93	0.9235	0.9121	0.9178
9	Xgboost	0.9638	0.9718	0.9427	0.957
10	LGBM	0.9648	0.9776	0.9394	0.9581
11	Catboost	0.9522	0.9552	0.932	0.9435
12	Extra Tree Classifier	0.9618	0.9725	0.9372	0.9546
13	ANN	0.9531	0.9575	0.9317	0.9445
14	MLP	0.9614	0.9732	0.9356	0.954
15	Stacking_ml	0.9657	0.9726	0.9467	0.9594
16	Stacking_ml_mlp	0.9657	0.974	0.9451	0.9593
17	Hard Voting	0.9648	0.9793	0.9377	0.958
18	**Soft Voting**	**0.966**	**0.9769**	**0.943**	**0.9596**

Figure 11: Evaluation Score for Models developed on First Dataset

Sr. No.	Technique	Accuracy	Precision	Recall	F1 Score
1	Random Forest	0.9281	0.9098	0.9504	0.9296
2	Decision Tree	0.9082	0.8865	0.9362	0.9107
3	Gradient Boosting Classifier	0.9294	0.907	0.9569	0.9313
4	KNN	0.5637	0.5341	0.9961	0.6954
5	Naive Bayes	0.9079	0.9119	0.903	0.9074
6	Logistic Regression	0.9289	0.9151	0.9455	0.93
7	SVMC	0.9594	0.98	0.9379	0.9585
8	Adaboost	0.9169	0.9022	0.9351	0.9184
9	Xgboost	0.9351	0.9268	0.9449	0.9358
10	LGBM	0.9349	0.92	0.9526	0.936
11	Catboost	0.8779	0.8537	0.9123	0.882
12	Extra Tree Classifier	0.9694	0.958	0.982	0.9698
13	ANN	0.69335	0.69338	0.69338	0.69338
14	MLP	0.9564	0.9297	0.9874	0.9577
15	Stacking_ml	0.9719	0.9656	0.9787	0.9721
16	**Stacking_ml_mlp**	**0.9738**	**0.9677**	**0.9803**	**0.974**
17	Hard Voting	0.9607	0.9568	0.9651	0.9609
18	Soft Voting	0.961	0.9514	0.9716	0.9614

Figure 12: Evaluation Score for Models developed on Second Dataset

4.2 Significance Testing

From the evaluation results, it can be seen that for the first dataset the classification models - Random Forest, Gradient Boosting, XgBoost, Light Gradient Boosting , Extra Tree Classifier, Multi-Layer Perceptron, Stacking models, Hard Voting and Soft Voting classifiers showcase a similar performance with slight differences of 1-3 percent. Similar circumstances can be observed in case of the second dataset. The classification models showcased a similar degree of performance with minor differences.

To check if the difference between the models is significant or not, significance testing is carried out (Divine et al.; 2013). The selection of significance test methods can be done by accessing the normal distribution of the variables. Post the normal distribution, it was observed that the evaluation scores are not normally distributed. Wilcoxon signed rank test is adopted when the variable is not normally distributed (Divine et al.; 2013). Hence Wilcoxon signed rank test was adopted in this study for accessing the significant difference between the performance of the models. The results of the significance test for the first and second dataset models can be seen that in the Figure 13 and Figure 14 respectively.

	Random Forest	Gradient Boosting	XgBoost	Light Gradient Boosting	Extra Tree Classifier	Multi Layer Perceptron	Stacking_1	Stacking_2	Hard Voting	Soft Voting
Random Forest	X	0.002	0.02	0.131	0.105	**0.002**	0.039	0.08	0.574	0.081
Gradient Boosting	0.504	X	0.798	0.08	0.102	**0.002**	0.304	0.386	0.557	0.117
XgBoost	0.383	0.798	X	0.178	0.28	**0.002**	0.383	0.505	0.644	0.256
Light Gradient Boosting	0.131	0.08	0.178	X	1	**0.002**	0.72	0.609	0.538	0.878
Extra Tree Classifier	0.105	0.102	0.28	1	X	**0.002**	0.959	0.959	0.327	0.877
Multi Layer Perceptron	**0.002**	**0.002**	**0.002**	**0.002**	**0.002**	X	0.002	0.002	0.002	0.002
Stacking_ml	0.039	0.304	0.383	0.72	0.959	**0.002**	X	0.836	0.682	0.959
Stacking_ml_mlp	0.08	0.386	0.505	0.609	0.959	**0.002**	0.836	X	0.474	0.798
Hard Voting	0.574	0.557	0.644	0.538	0.327	**0.002**	0.682	0.474	X	0.441
Soft Voting	0.081	0.117	0.256	0.878	0.877	**0.002**	0.959	0.798	0.441	X

Figure 13: Wilcoxon signed rank rest results for First Dataset

	Random Forest	Gradient Boosting	SVM	Logistic Regression	XGBoost	LightGBM	Extra Trees Classifier	MultiLayer Perceptron	Stacking_ml	Stacking_ml_mlp	Hard Voting	Soft Voting
Random Forest	X	0.438	0.178	0.223	0.188	0.062	0.312	0.438	0.625	0.59	0.062	0.062
Gradient Boosting	0.438	X	0.062	0.125	0.812	0.312	0.59	1	0.062	0.312	0.062	0.062
SVM	0.178	0.062	X	0.312	0.178	0.062	0.059	0.062	0.312	0.062	0.178	0.059
Logistic Regression	0.223	0.125	0.312	X	0.188	0.062	0.312	0.059	1	0.281	0.125	0.062
XGBoost	0.188	0.812	0.178	0.188	X	0.188	0.588	1	0.188	1	0.062	0.062
LightGBM	0.062	0.312	0.062	0.062	0.188	X	0.312	0.125	0.125	0.125	0.062	0.062
Extra Trees Classifier	0.312	0.59	0.059	0.312	0.588	0.312	X	1	0.062	0.625	0.062	0.062
MultiLayer Perceptron	0.438	1	0.062	0.059	1	0.125	1	X	0.312	0.625	0.062	0.062
Stacking_ml	0.625	0.062	0.312	1	0.188	0.125	0.062	0.312	X	0.312	0.062	0.062
Stacking_ml_mlp	0.59	0.312	0.062	0.281	1	0.125	0.625	0.625	0.312	X	0.062	0.062
Hard Voting	0.062	0.062	0.178	0.125	0.062	0.062	0.062	0.062	0.062	0.062	X	0.625
Soft Voting	0.062	0.062	0.059	0.062	0.062	0.062	0.062	0.062	0.062	0.062	0.625	X

Figure 14: Wilcoxon signed rank rest results for First Dataset

Thus, it can be interpreted that the difference is significant except in case of Multilayer Perceptron classifier for first dataset. For rest of the models this study fail to reject the null hypothesis that there are no significant differences between the models.

Similarly for the all the classifiers, this study fails to reject the null hypothesis that there are no significant differences between the models.

4.3 Important Feature

In addition to the performance scores, the importance coefficients of the variables was also obtained. Figure 15 showcases the important variables derived by the different classifiers.

From the Figure 15 it is evident that the as for service attributes, in-flight Wi-Fi services, online boarding provision and 'Type of Travel' are an important aspects of passenger's satisfaction. Hence, the airline crew should behave accordingly with the passenger's travelling for business or personal travel purposes.

Model	Important Features
Random Forest	Type_of_Travel','Class', 'Inflight_wifi_service', 'Online_boarding', 'Inflight_entertainment', 'Leg_room_service'
Decision Tree	Type_of_Travel', 'Inflight_wifi_service', 'Online_boarding'
Gradient Boosting	Type_of_Travel', 'Inflight_wifi_service', 'Online_boarding'
Logistic Regression	'Inflight_wifi_service', 'Online_boarding'
Adaboost	Age', 'Inflight_wifi_service', 'Online_boarding', 'Seat_comfort'
Xgboost	Type_of_Travel', 'Inflight_wifi_service', 'Online_boarding'
Extra Tree Classifier	Type_of_Travel', 'Inflight_wifi_service', 'Online_boarding'

Figure 15: Important Feature

5 Discussion, Conclusion and Future Work

This study investigated the satisfaction of airline passengers by leveraging data mining and machine learning techniques on two different datasets. Overall 18 different binary classifications were developed for investigating the satisfaction of airline passengers. Out of which, for the first dataset, the Soft Voting classifier outperformed all the other classifiers with an accuracy of 96.60 percent. In the case of second dataset, the stacking classifier outperformed the other models with an accuracy of 97.38 percent. By how much margin these two classifiers are better than the other candidate models is showcased in the Figure 18. Also, it was observed that these two classifiers outperformed the models proposed by the previous studies as showcased in Figure 16 and Figure 17. In addition, by how much percent are proposed classifiers better than others models can be seen in Figure 18. Although the proposed classifiers showcase higher performance scores, but they have one major inefficacy in terms of training time. The proposed classifiers take a lot of time for training. In addition to the quantitative analysis, this study also highlighted the important determinants of passenger satisfaction.

Ref	Year	(First Dataset) Approach	Accuracy
(Noviantoro and Huang; 2022)	2022	DL(NN)	93.46%
(Tan; 2021)	2021	RF	93.58%
(Gao et al.; 2021)	2021	RF	0.96%
This study	2022	Soft Voting Classifier	0.97%

Figure 16: Performance Comparison with previous studies for Dataset 1

Ref	Year	(Second Dataset) Approach	Accuracy
(Gupta et al.; 2021)	2021	Dropout Model	79.44%.
(Tusar and Islam; 2021)	2021	SVM and LR	77.00%
(Rane and Kumar; 2018)	2018	AdaBoost	84.50%
(Wan and Gao; 2015)	2015	Voting Classifier	Precesion - 91.7%
(Aljedaani et al.; 2022)	2022	DL (LSTM, LSTM-GRU)	97%
(Kalangi et al.; 2021)	2021	LSTM	73.46%
(Qudar and Mago; 2020)	2020	TweetBERTv2	92.99%
(Rustam et al.; 2019)	2019	Voting Classifier	0.79%
(Kamruzzaman et al.; 2021	2021	Single CNN + FastText	0.93%
This study	2022	Stacking	97.38%

Figure 17: Performance Comparison with previous studies for Dataset 2

The main findings of this study can be summarized as follows:

- Stacking and Voting Classifiers showcase better performance results in classifying the sentiments of airline passengers because of their ability to combine results of multiple classifiers.
- Although the stacking voting classifiers showcase better classifications results, it should be noted that these techniques take a lot of time for training. Hence usage of high computing resources is recommended.

- Important features that need to be focused by the airline companies to increase the passenger satisfaction are - in-flight Wi-Fi services, online boarding provision and staff behavior with the passengers traveling for business or personal travel purposes.

Technique	Accuracy
Random Forest	0.29%
Decision Tree	1.99%
Gradient Boosting Classifier	0.23%
KNN	3.89%
Naive Bayes	10.18%
Logistic Regression	9.35%
SVMC	1.08%
Adaboost	3.60%
Xgboost	0.22%
LGBM	0.12%
Catboost	1.38%
Extra Tree Classifier	0.42%
ANN	1.29%
MLP	0.46%
Stacking_ml	0.03%
Stacking_ml_mlp	0.03%
Hard Voting	0.12%

(i)First Dataset: Soft Voting Vs. Rest of the Models

Technique	Accuracy
Random Forest	4.57%
Decision Tree	6.56%
Gradient Boosting Classifier	4.44%
KNN	41.01%
Naive Bayes	6.59%
Logistic Regression	4.49%
SVMC	1.44%
Adaboost	5.69%
Xgboost	3.87%
LGBM	3.89%
Catboost	9.59%
Extra Tree Classifier	0.44%
ANN	28.04%
MLP	1.74%
Stacking_ml	0.19%
Hard Voting	1.31%
Soft Voting	1.28%

(ii)Second Dataset: Stacking_ml_mlp Vs. Rest of the Models

Figure 18: Accuracy Comparison

As for future work, apart from investigating the performance of the proposed models on different datasets, deep learning classifiers can be developed for comparing the performance of the classifiers. In addition, prediction models for multi-class classification can be explored.

Acknowledgement

This study gratefully acknowledges the guidance given by the project supervisor Mr. Abdul Razaaq. His continuous guidance and support have contributed notably in this work and have led to significant improvements.

References

Adeola, M. M. and Adebiyi, S. O. (2014). Service quality, perceived value and customer satisfaction as determinant of airline choice in nigeria, *International Letters of Social and Humanistic Sciences* **20**: 66–80.

Aljedaani, W., Rustam, F., Mkaouer, M. W., Ghallab, A., Rupapara, V., Washington, P. B., Lee, E. and Ashraf, I. (2022). Sentiment analysis on twitter data integrating textblob and deep learning models: the case of us airline industry, Knowledge-Based *Systems* **255**: 109780.

Atallah, R. and Al-Mousa, A. (2019). Heart disease detection using machine learning majority voting ensemble method, *2019 2nd international conference on new trends in computing sciences (ictcs)*, IEEE, pp. 1–6.

Ayele, W. Y. (2020). Adapting crisp-dm for idea mining: a data mining process for generating ideas using a textual dataset, *International Journal of Advanced Computer Sciences and Applications* **11**(6): 20–32.

Chang, Y.-H. and Yeh, C.-H. (2002). A survey analysis of service quality for domestic airlines, *European journal of operational research* **139**(1): 166–177.

Chonsalasin, D., Jomnonkwao, S., Chanpariyavatevong, K., Laphrom, W. and Ratanav araha, V. (2022). Modeling of airline passenger loyalty: A comparison of leisure and business travelers, *Research in Transportation Business & Management* **43**: 100735.

Divine, G., Norton, H. J., Hunt, R. and Dienemann, J. (2013). A review of analysis and sample size calculation considerations for wilcoxon tests, *Anesthesia & Analgesia* **117**(3): 699–710.

Fernandez, A., Garcia, S., Herrera, F. and Chawla, N. V. (2018). Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary, *Journal of artificial intelligence research* **61**: 863–905.

Gao, K., Yang, Y. and Qu, X. (2021). Examining nonlinear and interaction effects of multiple determinants on airline travel satisfaction, *Transportation Research Part D: Transport and Environment* **97**: 102957.

Gupta, M., Kumar, R., Walia, H. and Kaur, G. (2021). Airlines based twitter senti ment analysis using deep learning, *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*, IEEE, pp. 1–6.

Hasan, T., Matin, A. and Joy, M. S. R. (2020). Machine learning based automatic classification of customer sentiment, *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, IEEE, pp. 1–6.

Johar, M., Ab Yajid, M. S. and Azam, S. F. (2020). What policies airlines in malaysia adopt do to achieve competitiveness?, *Systematic Reviews in Pharmacy* **11**(1): 770–778.

Kalangi, R. R., Maloji, S., Tejasri, N., Chand, P. P. and Ch, V. P. (2021). Sentiment analysis using machine learning, *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, pp. 116–121.

Kamruzzaman, M., Bakchy, S. C., Hasan, T., Shovon, M. M. I. and Hossain, M. (2021). An intense investigation on deep and traditional machine learning approaches for better sentiment classification, *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)*, IEEE, pp. 1–6.

Khuriwal, N. and Mishra, N. (2018). Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm, 2018 IEEMA *engineer infinite conference (eTechNxT)*, IEEE, pp. 1–5.

Li, Y., Yang, Z., Chen, X., Yuan, H. and Liu, W. (2019). A stacking model using url and html features for phishing webpage detection, *Future Generation Computer Systems* **94**: 27–39.

Moro, S., Laureano, R. and Cortez, P. (2011). Using data mining for bank direct marketing: An application of the crisp-dm methodology.

Neshir, G., Rauber, A. and Atnafu, S. (2021). Meta-learner for amharic sentiment clas sification, *Applied Sciences* **11**(18): 8489.

Noviantoro, T. and Huang, J.-P. (2022). Investigating airline passenger satisfaction: Data mining method, *Research in Transportation Business & Management* **43**: 100726.

Pavlyshenko, B. (2018). Using stacking approaches for machine learning models, 2018 IEEE *Second International Conference on Data Stream Mining & Processing (DSMP)*, IEEE, pp. 255–258.

Punel, A., Hassan, L. A. H. and Ermagun, A. (2019). Variations in airline passenger expectation of service quality across the globe, *Tourism management* **75**: 491–508.

Qudar, M. M. A. and Mago, V. (2020). Tweetbert: a pretrained language representation model for twitter text analysis, *arXiv preprint arXiv*:2010.11091 .

RaE, C. L., Metz, M., Toadere, C. et al. (2022). The role of service quality in ensuring customer satisfaction in the airline industry, *Ovidius University Annals, Economic Sciences Series* **22**(1): 708–715.

Rane, A. and Kumar, A. (2018). Sentiment classification system of twitter data for us airline service analysis, 2018 IEEE *42nd Annual Computer Software and Applications Conference (COMPSAC)*, Vol. 1, IEEE, pp. 769–773.

Rustam, F., Ashraf, I., Mehmood, A., Ullah, S. and Choi, G. S. (2019). Tweets classific ation on the base of sentiments for us airline companies, *Entropy* **21**(11): 1078.

Schroer, C., Kruse, F. and Gomez, J. M. (2021). A systematic literature review on applying crisp-dm process model, *Procedia Computer Science* **181**: 526–534.

Sezgen, E., Mason, K. J. and Mayer, R. (2019). Voice of airline passenger: A text mining approach to understand customer satisfaction, *Journal of Air Transport Management* **77**: 65–74.

Sharma, P., Kumar, R. and Gupta, M. (2021). Impacts of customer feedback for online-offline shopping using machine learning, *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, IEEE, pp. 1696–1703.

Singh, B. (2021). Predicting airline passengers' loyalty using artificial neural network theory, *Journal of Air Transport Management* **94**: 102080.

Tan, C. (2021). Bidirectional lstm model in predicting satisfaction level of passengers on airline service, *2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, IEEE, pp. 525–531.

Tusar, M. T. H. K. and Islam, M. T. (2021). A comparative study of sentiment analysis using nlp and different machine learning techniques on us airline twitter data, *2021 International Conference on Electronics, Communications and Information Technology (ICECIT)*, IEEE, pp. 1–4.

Wan, Y. and Gao, Q. (2015). An ensemble sentiment classification system of twitter data for airline services analysis, 2015 *IEEE international conference on data mining workshop (ICDMW)*, IEEE, pp. 1318–1325.

Zhang, L., Yang, H., Wang, K., Bian, L. and Zhang, X. (2021). The impact of covid-19 on airline passenger travel behavior: An exploratory analysis on the chinese aviation market, *Journal of Air Transport Management* **95**: 102084.