Predicting Airline Passenger Satisfaction with Stacking Classifiers and Machine Learning Models

MSc Research Project
MSc in Data Analytics

Thanmayee Mandava
Student ID: x23204192

School of Computing
National College of Ireland

Supervisor: Abdul Qayum

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Thanmayee Mandava
Student ID:	x23204192
Programme:	MSc in Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Abdul Qayum
Submission Due Date:	29/01/2025
Project Title:	Predicting Airline Passenger Satisfaction with Stacking Classifiers and Machine Learning Models
Word Count:	7527
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:	28th January 2025

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):	

Predicting Airline Passenger Satisfaction with Stacking Classifiers and Machine Learning Models

Thanmayee Mandava
x23204192

Abstract

In the airline industry, the important aspect that impacts an airline's performance is the satisfaction of passengers. Airlines that can understand and satisfy their passengers' expectations succeed through customer loyalty and increased sales. In addition to ticket pricing, modern passengers assess the kind of services provided by various carriers before making their choices. Therefore, airlines need to ensure service excellence. Passenger feedback leads to these expectations, and consequently, the airlines will take note of these for fine-tuning. This research has performed data mining on a dataset of over 130,000 customer satisfaction ratings for various airlines to analyze the main drivers of satisfaction. Even with increasing competition, most studies in the past failed to address the complexity of passenger satisfaction, often limiting either the factors or traditional methods used. This study focused on bridging the gap that identifies and predicts satisfaction drivers by embedding stacking classifiers and machine-learning models. The best performance was obtained for the Stacking model with the meta-learner classifier MLP, with a 96.53% accuracy, 97.97% precision, 94.06% recall, 95.97% F1 score, and 92.99% MCC. The study has underpinned key fact-oriented decision-making in achieving high satisfaction of customers in the airline sector to gain a competitive advantage.

Keywords— Airline satisfaction, machine learning, stacking model, customer satisfaction, data mining.

1 Introduction

The airline industry has grown in competition, with airlines manufacturing their existence in the market through access and innovation. Over time, budget airline today, customers have more choices than ever, and passenger satisfaction has become the most important differentiator in the success of any airline company (Wu and Gao; 2024). Therefore, airlines that meet or exceed customer expectations retain loyal customers while simultaneously adding new ones due to improved service quality and operational excellence (Noviantoro and Huang; 2022). Anticipating passenger requirements is an airline's strategic priority to retain competitiveness within this dynamic industry (Dike et al.; 2024).

Passenger feedback is one of the key resources that can help you track customer preference and enhance service standards. Due to better data collection and analysis, airlines are now able to use this feedback to understand how satisfied their customers are. Other than conventional mechanisms of surveys, Twitter or Facebook is one of the famous spots where passengers get to air their opinions. Mining such data provides a good deal of value to airlines in terms of sentiment analysis and feedback in real-time, hence helping them to spot strengths and weaknesses. Indeed, machine learning and data mining techniques have been effective processes for such voluminous structured and unstructured data, hence leading to actionable insights (Park et al., 2022; Umar et al., 2023). This research examines the detailed analysis of passenger satisfaction data, including a survey dataset of ratings from passengers. The determination of the satisfaction variables will be analyzed using machine learning and ensemble methods. This helps to get insights into which services build loyalty while failures cause dissatisfaction.

(Bellizzi et al.; 2022) identified the key drivers of passenger satisfaction attributes like staff helpfulness and professionalism or how smooth the boarding process is. In their research, they employed a simple technique to evaluate passenger responses and pinpoint the important areas that can enhance customer experience for airlines. The study analysis breaks the data into easily understandable components to detect patterns and relationships. It provides clear outcomes, indicating to airlines which factors, whether service quality or boarding efficiency, have the most impact on passenger satisfaction. They also applied a ranking method to those factors to guide airlines on where to concentrate to realize improvements that are most valuable to passengers. Airlines would greatly benefit from this realistic approach, which avoids too intricate tactics, as it would help them improve their services.

Given the leveraging of advanced methodologies and integrating feedback from various sources, this research helps to provide useful insights for airlines to further enhance the quality of their service, thereby enhancing passenger satisfaction and retaining competitiveness in the dynamically changing aviation sector.

1.1 Research Question

How well do stacking ensemble models perform in increasing the reliability and predictive accuracy of passenger satisfaction in the airline industry?

1.2 Research Objectives

In addressing the above research question, the following goals have been noted:

1. Comprehensive analysis of the current techniques and algorithms being used for passenger satisfaction.
2. An extensive empirical analysis between machine learning models, neural networks, and ensemble classifiers.
3. Pinpointing the main features that greatly influence passenger satisfaction and need attention for improvement.

This research study is further divided as follows:

- Section 2 provides a briefing on a related study on airline passenger happiness.
- In Section 3, the Methodology used in the study is explained.
- Section 4 deals with the Design specification.
- Section 5 concentrates on the suggested implementation.
- Section 6 evaluates the performance of different models and their metrics.
- Finally, the study is concluded with the conclusion and future work in Section 7.

2 Related Work

Research in airline passenger satisfaction has expanded significantly over time, having drawn the attention of many researchers beyond simple surveys into more advanced data-driven methods that allow an understanding of what keeps consumers satisfied and loyal. Passenger needs became even more crucial with increased competition within the airline industry. Projects today use various sources of data, from survey responses to social media posts, to map the needs and experiences of passengers. Machine learning has, in particular, become popular, since it can yield more accurate predictions by using techniques like stacking and combining different models that can produce a more complex number of factors lying beneath passenger satisfaction.

The various literature pieces existing concerning airline passenger satisfaction can be identified into different approaches, reflecting the evolution and growth of experience of the analysis techniques in this field:

2.1 Traditional Models and Early Machine Learning Approaches

In (Shah et al.; 2020), the authors studied the service quality of airlines and the factors affecting behavioral intentions and ensuring customer loyalty through hierarchical regression on the survey data. The authors have pointed to customer satisfaction as the critical mediator between service quality and loyalty, with a focus on punctuality, in-flight service, and responsiveness of staff as key aspects (Song et al.; 2024). The authors in (Shah et al.; 2020) used regression analysis to predict customer satisfaction and achieved an adjusted R^2 value of 78%. However, this approach had few limitations in modeling nonlinear effects and interactions. This limitation in the study provides a way to use advanced machine learning techniques to provide better performance to predict passenger satisfaction.

Two studies, (Suprapto and Oetama; 2023) and (Nurdina and Puspita; 2023), looked into the use of machine learning techniques to predict airline passenger satisfaction. In one of the studies, (Suprapto and Oetama; 2023) compared Decision Tree and Naïve Bayes models using data from U.S. airlines. The findings from the study show that Naïve Bayes performed better, achieving 86.67% accuracy, 85.4% precision, 87.1% recall, and an 86.2% F1-score. The important features included business class, in-flight entertainment, and cleanliness. (Nurdina and Puspita; 2023) used the Kaggle dataset and were able to work with records of 26,000, reaching 84.48% accuracy in Naïve Bayes while the accuracy of KNN turned out to be 65.38% only. These studies have pointed out both the strengths of Naïve Bayes but at the same time consequences related to its inability to handle complex interdependencies. A stacking ensemble technique might combine the powers of existing models to improve predictions across high-dimensional satisfaction factors.

The Study (Hong et al.; 2023) used several machine-learning techniques to predict the satisfaction levels of passengers over a data set with 129k records. Random Forest recorded the highest performance among other algorithms, which were 89.20% accuracy, 93.04% precision, 84.92% recall, and 88.80% F1-score. The main factors of satisfaction included online check-in, in-flight entertainment, seat comfort, cleanliness, and In-flight Wi-Fi service. Random forest performed well with the high-dimensional data, however using just one ensemble model was a limitation since this did not reflect the existing heterogeneity of passenger satisfaction. This study highlights the ability of these stacking ensemble approaches used in synergism to maximize the power of different models for predictive accuracy. This research also has implications for passenger satisfaction and service quality in the competitive airline business.

Early machine-learning models were more flexible and accurate than classic statistical methods, but the simplicity of available algorithms imposed limits. While Decision Trees and Naïve Bayes were fast, they could not record complicated interdependencies. Improvements came from ensemble methods such as Random Forest but depending on any single model was limiting in terms of flexibility and robustness. This raises the advanced multi-layer approach, like the stacking tradition.

2.2 Using Unstructured Data like Text and Social Media

The passenger is shifting towards an expectation, and real-time sentiment analysis will have to be deployed to identify the drivers of satisfaction. Airlines are flooded with unstructured data, such as online reviews and social media, that can offer companies spontaneous feedback about where they might improve. However, the size and variability to use advanced machine learning techniques while analyzing this data to retrieve actionable insights using sentiment analysis as well as natural language processing (NLP).

The study (Hong et al.; 2023), examined the use of a Support Vector Machine classifier to predict sentiment from over 1.3 million tweets on ten major U.S. airlines. The proposed model obtained a 91.86% accuracy, 89.40% precision, 92.10% recall, and 90.20% F1-Score, validated through 10-fold cross-validation with an average accuracy of 90.03%. Time-series analysis was done using Bollinger Bands to track the trends in sentiment. These were compared with real-world events, such as flight cancellations or the sudden ending of the COVID-19 mask mandate, to see corresponding spikes or drops in sentiment.

One of the approaches, (Mirthipati; 2024) achieved 92.3% accuracy (91.1% precision, 92.6% recall, and 91.8% F1-score) by CNN model, outperforming other models such as Support Vector Machine (SVM) and Artificial Neural Networks (ANN). Using word embedding methods and association rule mining, (Mirthipati; 2024) revealed various negative sentiments such as delays in the flight, bad food quality, and lost baggage at a large scale, and positive sentiments such as awesome crew service. Although CNNs were good for modeling patterns in unstructured text data, the a lack of flexibility in incorporating heterogeneous data. The implication is that if CNNs are combined with other machine learning models like stacking, it could be more adaptable, as unstructured textual understanding can be integrated with the structured measure of satisfaction.

Another study conducted by, (Dike et al.; 2024), examined 17,726 consumer reviews at Skytrax using the SERVQUAL and Expectation Confirmation frameworks. Five aspects of service quality were identified: tangibles, certainty, responsiveness, empathy, and dependability. Empathy and reliability turned out to be significant predictors of satisfaction, explaining 36.2% of service quality perceptions. The CFA indicated a good fit with CFI = 0.92, and RMSEA = 0.03, with high reliability. It included seating, food quality, customer service, baggage handling, and delays. This workout showed the potential of stacking models in integrating qualitative and quantitative insights into becoming more data-driven in improving passenger satisfaction.

(Baydogan and Alatas; 2019) based on Natural Language Processing (NLP) analyze the customers of six airline companies and classify them into positive, neutral, or negative. For US Airways, the SMO classifier yielded the best performance at 79.7% with the precision, recall, and F1-score of 74.9%, 75.6%, and 75.2%, respectively. However, the Naïve Bayes classifier performed the best when ROC-AUC (77.5%) was taken into account. Dealing with imbalance was one of the takeaways highlighted by the study which had more negative than positive and neutral reviews. Thus, the models offered valuable insights into the sentiment patterns in their respective analyses but were significantly affected by feature interdependencies and data imbalance that might be improved with the use of a stacking framework to provide more consistency in prediction moving forward.

2.3 Using Machine Learning Approaches for Passenger Satisfaction

Machine learning might serve as a very strong tool for passenger satisfaction assessment in the rising competition and growing customer expectations by forecasting and analysis of customer feedback. As mentioned by (Dike et al.; 2024), unlike traditional methods, ML algorithms can handle high-dimensional data to uncover complex nonlinear relationships among satisfaction factors. It is the flexibility of ML approaches that allows them to be particularly effective in refining predictive models and drivers of customer loyalty by offering actionable insights for airlines.

In another study (Kumar et al.; 2024), the author used a dataset of 129k observations to predict airline passenger satisfaction was used (2024). The random forest model gave 94.2% accuracy and 92.7% precision, 93.5% recall as well a 93.1% F1-score, significantly increasing seat comfort, inflight entertainment as well and online booking features as the main determinants of customer satisfaction. Although performing well in high-dimensional data, random forests had issues with imbalances in datasets and dependencies of generated features. The study came up with recommendations to integrate Random Forest within a stacking framework that could offer higher predictive accuracy and be robust to utilize complementarities among models providing actionable insights to stakeholders, and improving passenger satisfaction and efficiency in operations.

The study (Salah-Ud-Din et al.; 2024), has predicted the satisfaction level-satisfied, neutral, or dissatisfied-using Gradient Boosting from a Kaggle dataset containing 100K records. The focus of the attention has fallen on Wi-Fi services, Boarding processes, and in-flight entertainment. Gradient Boosting yielded the best accuracy of 95.16%, with precision at 94.00%, recall at 94.50%, and equally an F1-score at 94.25%. Other models that yielded promising performances included Random Forest, Naïve Bayes, and Logistic Regression. Gradient Boosting demonstrated great performance, iteratively redoing itself to fit the exact errors within complex relationships of data. However, on its own, it could run a certain risk of overfitting and high computation costs. It might be balanced, embedding it into a stacking model to combine the accuracy of this model with the efficiency of simpler ones for a more balanced approach to outcome prediction. In-flight Wi-Fi and the online boarding process were important passenger satisfaction factors that provided actionable insights into service enhancements.

Indeed, these approaches have led to a considerable increase in the precision and flexibility of models for predicting or classifying satisfaction. Individually unique strengths of algorithms result from limitations related to the use of single models and imply advanced ensemble techniques, such as stacking. Combining the outputs of multiple algorithms in the form of stacking models gives a more effective structure for gaining a broader range of satisfaction characteristics and relationships, allowing a stronger alternative basis for airline satisfaction models to emerge.

2.4 New Approaches with Stacking Techniques

In the current development of studies on airline customer satisfaction, there is an increasing demand for advanced predictive models that capture complex and multi-layered interactions within high-dimensional datasets. Classic machine learning models are sometimes effective, yet too limited to comprehensively address the nuanced relationships present in satisfaction data. An ensemble learning method called stacking combines multiple machine-learning models into a single model. Recently, applications of the stacking ensemble method to improve performance by exploiting complementary strengths of different algorithms have been explored as a very promising method for achieving high predictive accuracy and model robustness.

Stacking has been one of the promising approaches in airline satisfaction research to transcend some pitfalls of relying on single algorithms. Ensemble models such as Naïve Bayes combined with Random Forest can exploit a broad class of problems with high-dimensional accuracy on the efficiency of Naïve Bayes.(Salah-Ud-Din et al.; 2024), presents evidence that the combination of Gradient Boosting and simpler classifiers improves their accuracy in structured datasets. On the other hand, this work also comments on computational issues of Gradient Boosting, stacking can mitigate this by adding simpler and faster models to balance the load without losing accuracy.

(Wu and Gao; 2024) investigated stacking models for airline passenger satisfaction prediction using a set of more than 100K reviews . As far as the stacking, it performed much better than any of the traditional ensemble methods, such as Boosting and Bagging, with 95.8% accuracy and 94.7% F1-score. Some of the key satisfaction drivers that have been identified include check-in processes, flight punctuality, and customer support. The study showed that stacking is a robust approach to process noisy and imbalanced datasets, which are challenges in satisfaction analysis, also showing how it can improve predictive accuracy and model reliability.

Ensemble methods have been investigated in airline satisfaction prediction during the post-COVID-19 by (Lee; 2024), where various techniques, including Decision Tree, Random Forest, Boosting, and Stacking were compared on a diversified data set. Stacking represented the best performance model that achieved an accuracy of 96.26%, which follows Bagging with 96.12%, and Boosting at a rate of 96.04%. (Lee; 2024) focused on the fact that Stacking can combine the benefits from many strong base models and will make it very powerful when coping with high-dimensional and high-interdependent data. While stacking showed better predictive performance and robustness, the study expressed the need for further exploration with larger datasets and alternative stacking designs to realize its full potential. This approach goes in line with the trend of using advanced ensemble methods to address the evolving complexities of passenger satisfaction dynamics.

Stacking approaches have emerged as one of the key mechanisms for passenger satisfaction analysis by airlines while addressing the challenges that emanate from the conventional single-model machine learning and single-algorithm ensemble methods. This gives high performance, robustness, and the ability to manage high-dimensional and interdependent data by availing the model with complementary strengths of several models. This is the flexibility that makes it so key in capturing complex dynamics of satisfaction and providing actionable insights for customer experience improvement. Of course, further exploration involving larger datasets and the usage of alternative stacking designs remains to be done before it fully meets the evolving challenges facing the performance of satisfaction analysis.

Reference	Focus	Key Findings	Limitations	Techniques/Models Used	Dataset Used
(Shah et al., 2020)	Analyzed airline service quality and customer loyalty using regression.	Customer satisfaction mediates service quality and loyalty (Adjusted R^2 = 78%).	Unable to model non-linear effects and interactions.	Hierarchical Regression	Survey data from airline passengers
(Suprapto and Oetama, 2023)	Compared Decision Tree and Naïve Bayes for satisfaction prediction.	Naïve Bayes outperformed Decision Tree (86.67% accuracy, F1-score 86.2%).	Naïve Bayes struggles with complex interdependencies.	Decision Tree, Naïve Bayes	U.S. airline satisfaction data
(Nurdina and Puspita, 2023)	Examined Naïve Bayes and KNN on satisfaction datasets.	Naïve Bayes achieved 84.48% accuracy, but KNN lagged at 65.38%.	KNN showed poor performance; limited generalizability.	Naïve Bayes, KNN	Kaggle dataset with 26,000 observations
(Hong et al., 2023)	Applied ML models; Random Forest performed best on large datasets.	Random Forest achieved 89.20% accuracy; key factors: online check-in, seat comfort, cleanliness.	Random Forest lacked heterogeneity representation.	Random Forest, SVM, Logistic Regression	Airline passenger satisfaction dataset (129,880 records)
(Mirthipati, 2024)	Used CNN for sentiment analysis on unstructured airline data.	CNN achieved 92.3% accuracy; identified sentiments like delays and crew service.	CNN struggled to incorporate structured data.	CNN, ANN, SVM	Airline passenger tweets
(Dike et al., 2024)	Used SERVQUAL framework to study satisfaction dimensions.	Empathy and reliability significantly predicted satisfaction (CFI = 0.92, RMSEA = 0.03).	Study focused on specific dimensions; broader integration required.	SERVQUAL Framework	Skytrax customer reviews (17,726 observations)
(Baydogan and Alatas, 2019)	Analyzed sentiment using NLP and machine learning classifiers.	SMO classifier had 79.7% accuracy; Naïve Bayes performed best on ROC-AUC (77.5%).	Data imbalance and feature interdependencies affected performance.	NLP, SMO Classifier, Naïve Bayes	Sentiment data from six airlines
(Salah-Ud-Din et al., 2024)	Predicted satisfaction using Gradient Boosting on Kaggle datasets.	Gradient Boosting achieved 95.16% accuracy; key drivers: Wi-Fi, boarding process.	High computation cost and overfitting risks in Gradient Boosting.	Gradient Boosting, Random Forest, Naïve Bayes	Kaggle dataset with 100,000 records
(Wu and Gao, 2024)	Investigated stacking models for satisfaction prediction.	Stacking outperformed other methods (95.8% accuracy, F1-score 94.7%).	Needs exploration with larger datasets and alternative stacking designs.	Stacking, Boosting, Bagging	Airline passenger reviews (100,000+ records)
(Lee, 2024)	Compared ensemble techniques; stacking performed best in prediction.	Stacking achieved highest accuracy (96.26%) compared to Boosting (96.04%) and Bagging (96.12%).	Requires further testing on larger datasets for full potential.	Stacking, Boosting, Bagging	Post-COVID airline satisfaction dataset
(Song et al., 2024)	Explored customer satisfaction as a mediator between service quality and loyalty.	Customer satisfaction was a critical mediator (Adjusted R^2 = 78%) between service quality and loyalty.	Did not account for non-linear effects or interactions.	Hierarchical Regression	Customer survey data on service quality
(Kumar et al., 2024)	Analyzed various ML models for predicting passenger satisfaction using 129,880 records.	Random Forest achieved 94.2% accuracy, precision of 92.7%, recall of 93.5%, and F1-score of 93.1%.	Struggled with imbalanced datasets and feature interdependencies.	Random Forest, Stacking Framework	Airline passenger satisfaction dataset (129,880 records)

Figure 1: Key Findings and Techniques in Airline Satisfaction Studies

Figure 1 summarizes the findings from various studies on airline passenger satisfaction with a focus on findings, limitations, and approaches adopted. These various studies have various aspects while looking into aspects related to satisfaction factor identification, analysis, customer feedback improvement, prediction improvement, the dataset used, etc. Overall, their results indicate very relevant drivers like service quality, customer experience, and operational factor-based drivers. However, the studies have limitations of their own, such as managing complex relationships among data, experiencing imbalanced datasets, or accounting for subtle passenger conduct. Different methods are utilized in each study to overcome these hurdles, thus feeding vital directions into how to increase levels of satisfaction and provide practical approaches to improvements within the airline service sector.

3 Methodology

This research leverages CRISP-DM (Cross Industry Standard Process for Data Mining) for Data Mining to evaluate machine learning models of airline passenger satisfaction based on structured data. (Wu and Gao; 2024) applied this framework to examine a business objective to enhance customers' satisfaction, understanding of data, preparation, modeling, evaluation, and deployment.

Figure 2 gives a visual view of the methodology, showing steps from beginning to end. Firstly, the data collection is carried out, followed by its preprocessing steps like handling missing values, treatment of null values, encoding of categorical variables, going further with exploratory data analysis, and further data sampling to separate the data into test and training sets. The machine learning and deep learning models, along with stacking, have been developed on these train and test data sets.

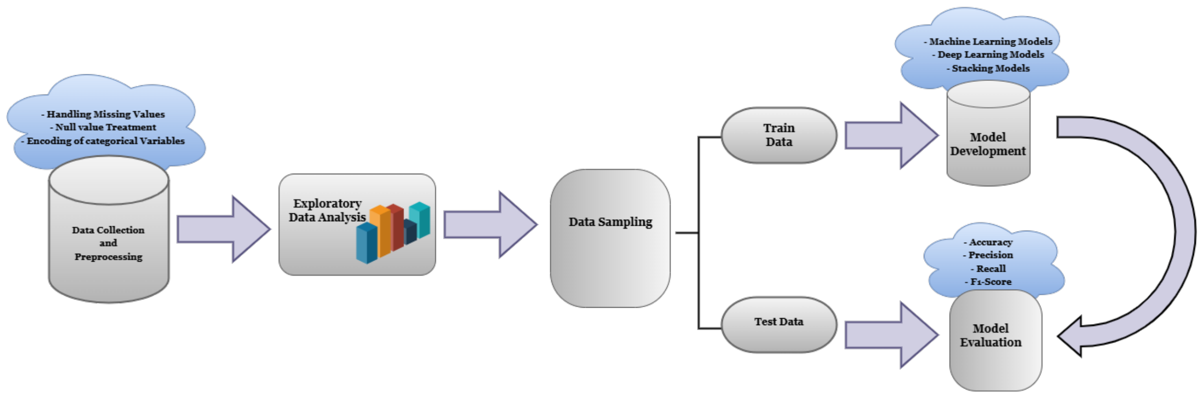

Figure 2: CRISP-DM Methodology

3.1 Data Understanding

The dataset used in this research is taken from the Kaggle [1] contains 129,880 entries and has all relevant information about airline passenger satisfaction(Baydogan and Alatas; 2019). Figure 3 describes all the features in the dataset like passenger description, details of travel, ratings related to services, and several operational factors such as baggage handling and flight delays. This research's target variable is satisfaction (whether it be neutral, dissatisfied, or satisfied).

Feature	Description
Id	A unique identifier for each passenger to prevent data duplication.
Gender	The passenger's gender, categorized as male or female.
Customer Type	Indicates whether the passenger is a Loyal Customer (repeat flyer) or a Disloyal Customer (infrequent flyer).
Age	The age of the passenger, represented as a numeric value.
Travel Type	Specifies whether the travel is for Business or Personal purposes.
Class	The travel class of the passenger, which can be Business, Eco Plus, or Economy.
Flight Distance	The distance traveled on the flight route, measured in miles.
Inflight Wi-Fi Service Rating	A score reflecting the quality and reliability of inflight Wi-Fi.
Convenience of Departure/Arrival Times	An evaluation of how suitable the departure and arrival times are.
Ease of Online Booking	Assesses how easy and user-friendly the online booking process is.
Gate Location	Indicates whether the boarding gate's location was convenient.
Food and Beverages	An assessment of the quality of food and beverage services provided onboard.
Online Boarding	A rating of the boarding process as experienced by passengers.
Seat Comfort	A score reflecting the comfort of the seating arrangement.
Inflight Entertainment	Availability and quality of inflight entertainment options.
Inflight Services	A general rating of the onboard personnel's services before and during the flight.
Leg Room	An evaluation of the adequacy of legroom space.
Baggage Handling	A score assessing the efficiency and performance of the baggage handling process.
Check-in Service	A rating of the check-in experience.
Inflight Service	A general satisfaction score for inflight services.
Cleanliness	A rating of the aircraft's cleanliness.
Departure Delay	The number of minutes the flight was delayed at departure.
Arrival Delay	The number of minutes the flight was delayed upon arrival.

Figure 3: Features and their description in the dataset

[1] https://www.kaggle.com/datasets/mysarahmadbhat/airline-passenger-satisfaction

3.2 Data Preparation

Data preparation is the step where raw datasets are converted into structured formats suitable for analysis and modeling. Data cleaning (missing value treatment, deleting duplicates, removing outliers and inconsistencies, etc.). Categorical variables including Customer Type, Gender, Class, and Type of Travel are transformed into numerical data by applying methods such as one-hot or label encoding (Dike et al.; 2024). Numerical variables like Flight Distance and satisfaction ratings are standardized or normalized so that they contribute equally to the performance of the model(Noviantoro and Huang; 2022).

The preprocessing step includes important steps to prepare it for machine learning analysis. Missing values were observed in the 'Arrival delay' column and were replaced with the median. The irrelevant column 'ID' was dropped to eliminate noise and reduce dimensionality. LabelEncoder is used to convert categorical variables into a numerical format to be compatible with machine learning algorithms. In model training, to improve consistency and prevent scale-related biases, numerical features were standardized to a standard scale. Furthermore, outliers detected in the Departure Delay column were removed using the IQR method to avoid the influence of extreme values, enhancing the robustness and accuracy of subsequent predictive modeling. These preprocessing steps were important in ensuring the dataset's reliability and effectiveness for further analysis.

3.2.1 Exploratory Data Insights

The data presents some important satisfaction patterns through variables and Figure 4 shows the average passenger satisfaction scores for various attributes.

1. 56.55% of the passengers are neutral or dissatisfied, while 43.45% are Satisfied, which indicates ample scope for service improvement.
2. Business Class passengers are quite satisfied, the neutral or dissatisfied rate in Economy Class is relatively high, and results from Economy Plus, although somewhat better, were also lagging.
3. The trends are similar, though females exhibit slightly higher levels of satisfaction.
4. It can be observed that the returning customers are more satisfied, while the first-time flyers show more neutral or dissatisfied responses owing to unfamiliarity with the services.
5. Business travelers are more satisfied, while personal travelers are less satisfied due to the unmet expectations over cost and comfort.
6. Meanwhile, passengers in the 20–40 years age group reported more balanced satisfaction, though the younger and older counterparts still remained highly dissatisfied because of diverging priorities over affordability, entertainment, or comfort.

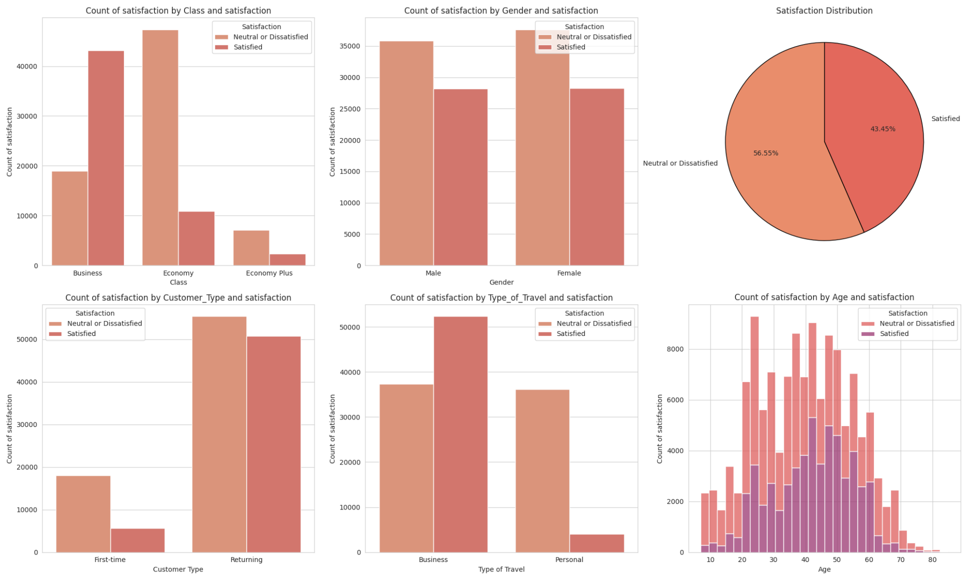

Figure 4: EDA steps for some of variables

In Figure 5, bar graphs highlight the different service attributes that relate to airline passenger satisfaction. Attributes related to Check-in Service, Online Booking, and Online Boarding are highly satisfied with most ratings at 4 or 5. While Seat Comfort, Gate Location, and Food & Drink have received an average rating, in-flight Wi-Fi and entertainment show dispersed ratings across the scale. The analysis underlines how online processes and in-flight services represent the most relevant drivers of passenger satisfaction.

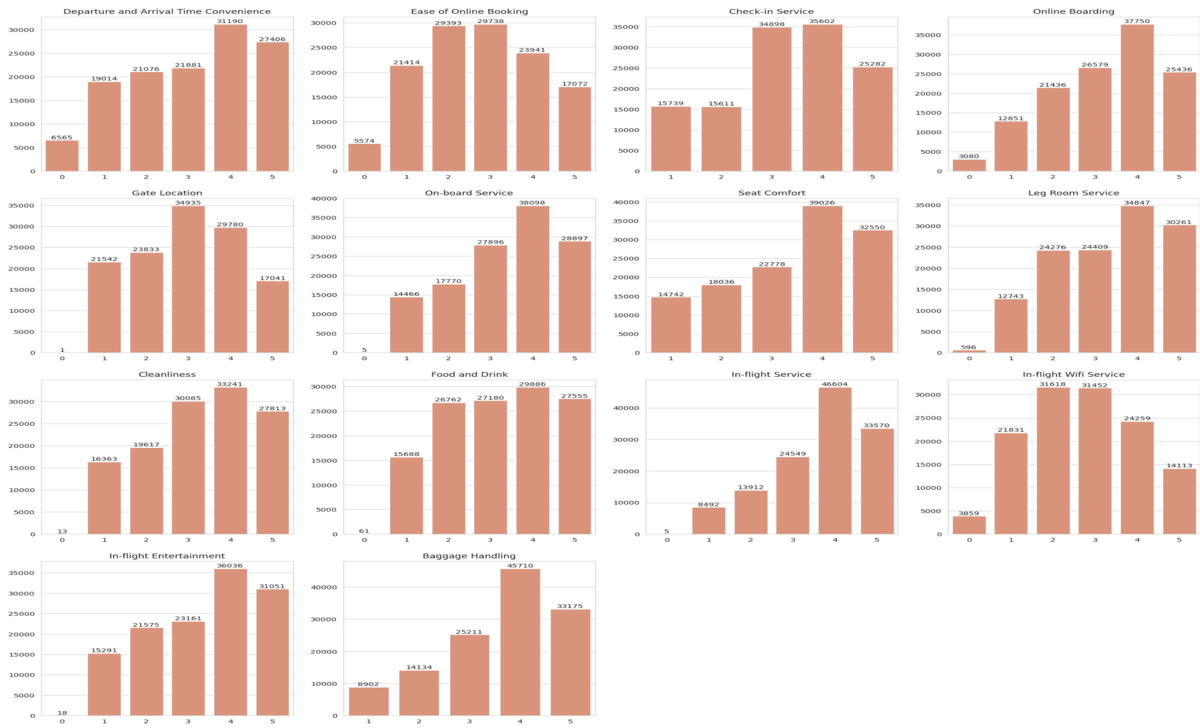

Figure 5: Passenger Ratings Across Service Attributes

3.3 Feature Engineering

Feature extraction transforms the attributes of the datasets and highlights the features that are strong predictors that enhance model performance. The most relevant features such as Inflight Wi-Fi Service, Seat Comfort, Online Boarding, and Flight Distance are detected while the categorical variables like Class, Gender, and Customer Type are numerically encoded (Hong et al.; 2023).

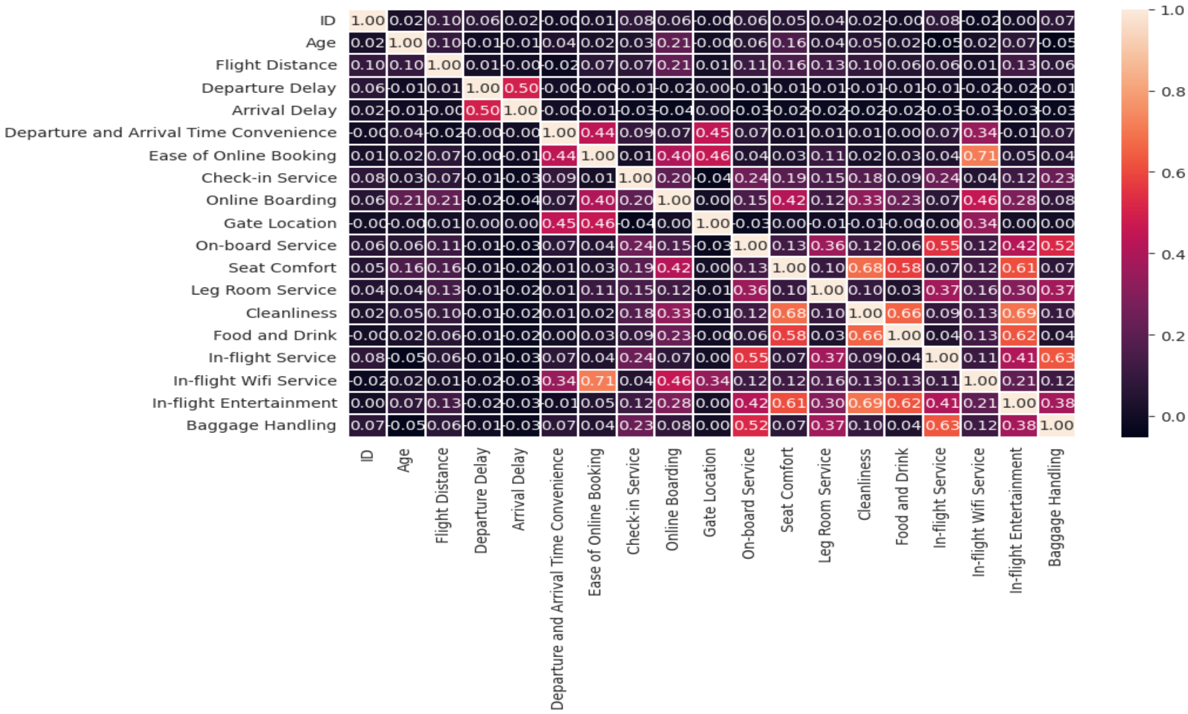

Figure 6: Correlation Matrix

Figure 6 represents the matrix of coefficients to visualize the relationship of variables in the airline satisfaction dataset, using color intensity to represent different strengths of correlations. It helps to identify key interactions and therefore potential predictors for passenger satisfaction.

1. The strong correlation (0.71) between Ease of Online Booking and In-flight WiFi Service reflects shared customer expectations in convenience and technological sophistication. That is passengers who care about seamless booking experiences also would expect high-quality inflight WiFi. Sometimes, airlines that invest a lot in seamless user experiences for booking systems migrate the same technological advancements into their other services to improve overall customer satisfaction.

2. A positive correlation (0.69) between in-flight entertainment and cleanliness shows that a clean cabin and the passenger's experience with entertainment, significantly influence overall satisfaction.

3. A positive correlation (0.68) between cleanliness and seat comfort means a passenger seeing a clean cabin is more likely to rate seat comfort positively. It leverages a sense of cleanliness which helps elevate the overall comfort of the experience making it more enjoyable.

4. The positive correlation (0.66) would suggest that the food and drink item quality is regarding cleanliness standards. Passengers associate clean dining areas with hygienic practices equating to better food quality, hence influencing their overall satisfaction with meals.

5. The correlation between in-flight service and baggage handling (0.63) highlights their importance to passengers when traveling.

3.4 Model Sampling

In Sampling, the dataset is divided into train and test sets, which are then used to develop machine-learning models. The dataset is divided into two parts for this study with a ratio of 80:20. 80% of the data is used for training and constructing the classification models, allowing the algorithms to learn from patterns within the data(Kumar et al.; 2024). The remaining 20%, on its part, was reserved for testing to measure the performance of their models on an unseen test set (Wu and Gao; 2024). This is an important separation in order not to let the models overfit and be generalizable on new data (Bellizzi et al.; 2022). The sampling strategy provided a reasonable basis to assess how well those models predicted passenger satisfaction based on input data and not using the same data for training and testing phases.

3.5 Model Development

During any data mining project, model development is an important phase, and different predictive and classification techniques are tried based on the research question. Many ML models are built in this stage and tested for their capability to handle the problem statement. For this dataset, various classification models were designed and tested. Figure 7 represent the details of these and the model parameters and configurations for the dataset. A detailed explanation of model development is given in Section 5.

S No.	Technique	Parameters
1	Random Forest	max_depth=24, random_state=42
2	Decision Tree	random_state=42
3	Gradient Boosting Classifier	estimators=100,max_depth=8, random_state=42
4	KNN	n_neighbors=15
5	Naive Bayes	default parameters
6	Logistic Regression	penalty='l2',random_state=42
7	SVM	kernel = 'rbf',gamma='auto',probability=True
8	AdaBoost	n_estimators=125, random_state=42
9	XGBoost	n_estimators=1000,max_depth=5
10	LightGBM	objective="binary", n_estimators=100
11	Catboost	iterations=50, depth=8, learning_rate=0.1, loss_function='RMSE'
12	Extra Tress Classifier	n_estimators=100, 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)
14	MultiLayer Perceptron	hidden_layer_sizes=(24,14),activation = 'relu', random_state=5, verbose=True,learning_rate_init=0.01
15	Stacking_ml	Base Classifiers - Random Forest, XgBoost,GBM, Meta Classifier - Logistic Regression
16	Stacking_ml_2	Base Classifiers - Random Forest,Extra Tress Classifier,GBM Meta Classifier - Logistic Regression
17	Stacking_ml_3	Base Classifiers - Random Forest,Extra Tress Classifier,LGBM Meta Classifier - Logistic Regression
18	Stacking_ml_4	Base Classifiers - Random Forest,XgBoost, Extra Tress Classifier Meta Classifier - Logistic Regression
19	Stacking_ml_mlp	Base Classifiers - Random Forest,XgBoost, GBM Meta Classifier - MLP
20	Stacking_ml_mlp_2	Base Classifiers - Random Forest, Extra Tress Classifier, GBM Meta Classifier -MLP
21	Stacking_ml_mlp_3	Base Classifiers - Random Forest, Extra Tree Classifier, LGBM Meta Classifier - MLP
22	Stacking_ml_mlp_4	Base Classifiers - Random Forest, Extra Tress Classifier, XGB Meta Classifier -MLP

Figure 7: Parameters used for ML models

3.6 Model Evaluation

Model evaluation is an important phase in evaluating the quality of machine learning models (Kumar et al.; 2024). In this study, the models were evaluated using metrics such as accuracy, precision, F1 score, recall, and MCC, which are standard measures for determining a model's predictive performance. While accuracy assesses overall predictive correctness, precision, and recall give us an idea about the quality of positive predictions to find all relevant instances. F1 score is the harmonic mean of precision and recall, combining the two types of metrics into one measure. These metrics are used to find the optimal model that provides unique and generalizable performance on unseen data in predicting passenger satisfaction (Hwang et al.; 2020).

4 Design Specification

Traditional approaches and older machine-learning models often do not capture the complexity of airline passenger satisfaction. So far, Models such as Decision Trees and Logistic Regression have shown meaningful insights but not only their ability to capture the non-linear relationship of features is weak but they also fail to consider interactions among some important variables such as Distance traveled by air, In-flight Service and Passenger related characteristics (Hong et al.; 2023). While other ensemble techniques showed better performance, such as the Random Forest Model and Gradient Boosting, reliance on these methods limits handling high-dimensional and heterogeneous data. Thus, advanced ensemble approaches have been necessitated with the integration of multiple algorithms for complete insights (Baydogan and Alatas; 2019).

In this study, a hybrid approach using ensemble learning techniques combined with advanced feature engineering and model evaluation strategies is suggested. We used stacking to combine the strengths of base models like Random forest, Gradient Boosting, and Decision Trees to ensure strong performance on all satisfaction indicators. Also, feature selection techniques will help us understand some important drivers influencing the satisfaction of passengers in the airline sector and improve model interpretability and efficiency.

Some of the metrics that will be used with the proposed framework are accuracy, precision, recall, F1-score, and MCC to ensure that the results obtained will be reliable. Results will be presented using visualization tools such as Plotly for a clear explanation to airline stakeholders regarding the trend of satisfaction and areas of improvement. With this, besides complementing the weaknesses of previous models, it scales and offers a practical solution for the service quality improvement of increasing customer satisfaction in a competitive environment such as the airline industry.

5 Implementation

In this research, various airline passenger satisfaction models have been considered, including Random Forest, Decision Tree, Gradient Boosting, K-Nearest Neighbor, Gaussian Naive Bayes, Logistic Regression, Support Vector Machine (SVM), AdaBoost, XGBoost, LightGBM, CatBoost, and Extra Trees.

The Random Forest Classifier has been instantiated with a maximum depth of 24 as a balance between the complexity of the model and the problem of overfitting. The Decision Tree model is developed using default parameters. For the Gradient Boosting model, 100 estimators with a maximum depth of 8 are used which are complex enough to find non-straightforward relationships between attributes without overfitting. The K-Nearest Neighbors model was tuned using 15 neighbors to reach an optimal balance between local generalization and noise reduction. Gaussian Naive Bayes and Logistic Regression baselines perform linearly by their respective default parameters.

Now, advanced ensemble models were further tuned for improved predictive accuracy. While AdaBoost used 125 estimators, the XGBoost utilized 1000 estimators with a maximum depth of 5 to effectively facilitate feature learning. LightGBM is a gradient-boosting model using 100 estimators with a binary objective which is used to handle the classification task quite well, while Extra Trees consisted of 100 decision trees for the sake of consistency. Lastly, the fitting of CatBoost should be done because it is designed for categorical data, in 50 iterations, with an RMSE loss function, to accommodate the structure of the dataset and arrive fast at the optimal solution.

These parameter values were chosen based on the recommendations in the literature, experimentation, and specific properties of the dataset with an attempt at a balanced computational cost predictive performance. This method made certain that each model used its respective strength and would identify the complex patterns present in the dataset.

6 Evaluation And Results

Different machine learning algorithms, including a stacking ensemble with an MLPClassifier as a meta-learner, are analyzed in this section. Each model was trained, validated, and tested on the passenger satisfaction dataset by comparing the accuracy, precision, recall, F1-score, and MCC in that order. The experimental framework keeps hyperparameters consistent for all models while systematically investigating the different configurations and their respective performances. Initial experiments considered the stacking classifier with all features of this dataset. Then, individual feature groups-including demographic information, flight details, or satisfaction-related service ratings-were analyzed in the next experiments. Finally, further experimentation was done on the impact caused by different combinations of feature subsets, including more base learners to diversify and improve model performance. These results gave insight not only into feature importance but also effectively compared the stacking classifier with other individual and ensemble models.

6.1 Experiment 1 – Baseline Model Performance

Base models showed variations in performance regarding the complexity of handling the airline passenger satisfaction dataset. The SVM model showed a good performance with 95.50% accuracy, and 94.82% F1-score, while the Random Forest, based on the ensemble structure, came out very efficient in handling overfitting to give 96.23% accuracy and a 95.65% F1-score. We can further observe that there are a couple more middle-range performances from tree-based models, such as Decision Tree with an accuracy of 94.65% and KNN with 92.85% accuracy. However, the simpler models like Naive Bayes with an accuracy of 86.47%, and Logistic Regression with 87.31% accuracy, didn't perform as well because of their limited capacity to capture complex relationships. Base models were reliable but less effective when compared to ensemble techniques.

6.2 Experiment 2 – Performance of Ensemble Models

Ensemble models performed the best, combining the prediction of multiple base classifiers. Among them, GBM and LGBM yielded the best results, providing an accuracy of 96.48% and 96.35%, respectively with 95.93% and 95.77% F1 scores, 92.88%. XGBoost runs very close at an accuracy of 96.17% with a 95.60% F1-score, while CatBoost has an accuracy of 95.38% and a 94.69% F1-score. Extra Trees have also reached high results with an accuracy of 96.16% and 95.56% of the F1-score. Gradient-boosting techniques emerged as the strongest, having the least number of errors and making pretty strong predictions on complex datasets.

6.3 Experiment 3 – Enhancing Predictive Accuracy with ANN and MLP

The results in Artificial Neural Network (ANN) and Multi-Layer perceptron (MLP) performance showcase the power of neural networks in classification problems. Performance metrics of the standalone ANN model such as accuracy, recall, precision, and F1-score were equal to 95.53%, 95.36%, 94.35%, and 94.85%, respectively, indicating their potential to record underlying relationships in the data. However, its lower recall suggests more limitations in detecting all the positive instances.

When optimized as a meta-learner within the stacking ensemble, the MLP substantially outperformed the stand-alone ANN. The stacking model with the three hidden layer improvement, dropout regularization, and early stopping of the MLP had the highest score at 96.17% accuracy, 96.15% precision, 95.03% recall, and 95.58% F1-score. This proves how much more powerful it can get when neural networks are combined with ensembling to enhance predictive performance and generalization.

6.4 Experiment 4 – Performance of Stacking Models

The stacking models implemented with a meta-learner Logistic Regression performed well on all evaluation metrics. Among all these models, the Logistic Regression stacking model achieved an accuracy ranging from 96.50%, whereas the best F1-score was 0.9597 in stack_model_ml. These models did a great job of balancing recall and precision, which can be observed from their MCC values Figure 8. Among them, the stack_model_ml outperformed other models with the highest precision of 97.27 % and an F1-score of 0.9597.

For MLP-based stacking models, stack_model_mlp_2 performed better in some configurations, within the accuracy of 96.53%, recording the best F1-score of 95.98%. This model also recorded the highest precision of 97.97% and MCC value of 92.99% among all variations, showing how good the MLP meta-learner is in the integration of base learner predictions. In summary, although all the meta-learners reached highly competitive results, stack_model_mlp_2 surpassed the leading stacking model by perfectly balancing the precision, recall, and overall generalization that led to an optimal classification, which made it perfect for predicting airline passenger satisfaction predictions.

	Model	Accuracy	Precision	Recall	F1 Score	MCC
0	Random Forest	0.962388	0.972662	0.940893	0.956514	0.92379
1	Decision Tree	0.946528	0.936852	0.941856	0.939348	0.891546
2	GBM	0.964891	0.975906	0.943433	0.959394	0.928899
3	KNN	0.928588	0.956128	0.877846	0.915316	0.856102
4	Naive Bayes	0.86476	0.868144	0.816375	0.841464	0.724784
5	Logistic regression	0.873114	0.873208	0.832224	0.852224	0.741823
6	SVM	0.955074	0.960395	0.936427	0.94826	0.9088
7	AdaBoost	0.929281	0.926176	0.911821	0.918943	0.856312
8	XgBoost	0.961772	0.965616	0.94676	0.956095	0.922391
9	LGBM	0.963543	0.976782	0.939405	0.957729	0.926246
10	Cat Boost	0.953881	0.958303	0.935814	0.946925	0.906359
11	Extra Tree	0.961695	0.97339	0.938529	0.955642	0.922431
12	ANN	0.955362	0.95363	0.943573	0.948575	0.909184
13	MLP	0.961734	0.961514	0.950334	0.955891	0.922152
14	Stacking Classifier	0.965045	0.972662	0.94711	0.959716	0.929111
15	stack_model_ml_2	0.964891	0.973078	0.946322	0.959513	0.928812
16	stack_model_ml_3	0.964313	0.97236	0.945709	0.958849	0.927635
17	stack_model_ml_4	0.963928	0.969626	0.947636	0.958505	0.926801
18	stack_model_mlp	0.965083	0.979214	0.940543	0.959489	0.929413
19	stack_model_mlp_2	0.965353	0.979752	0.94063	0.959793	0.929972
20	stack_model_mlp_3	0.964544	0.973397	0.945184	0.959083	0.928127
21	stack_model_mlp_4	0.965006	0.976257	0.943345	0.959519	0.929142

Figure 8: Results of all models

6.5 Critical Features Influencing Passenger Satisfaction

Figure 9 compares the important features of different models developed using machine learning. Online Boarding, Type of travel, and In-flight Wi-Fi Service were among the most consistently important predictors across models, and thus universally important for any satisfaction analysis. Other features in the model, such as Seat Comfort, Age, and Legroom Service, though specific to the model, have a certain relevance. The important observation to be made from this comparison is the strength of Online Boarding and Inflight Wi-Fi Services as reliable indicators across multiple algorithms, hence helping in the generalization and interpretability of models.

Machine Leaning Model	Important Features
Random Forest	Online Boarding, Inflight Wifi Service, Class, Type of Travel, Inflight Entertainment
Decision Tree	Online Boarding, Inflight Wifi Service, Type of Travel
Gradient Boost Classifier	Online Boarding, Inflight Wifi Service, Type of Travel
Adaboost	Inflight Wifi Service, Seat Comfort, Age, Online Boarding, Legroom Service
Xgboost	Online Boarding, Type of Travel, Inflight Wifi Service
Light GBM	Inflight Wifi Service, Age, Customer Type
Extra Tree Classifier	Type of Travel, Inflight Wifi Service, Online Boarding, Class

Figure 9: Crucial features for different ML models

Figure 10 explains the main factors of airline passenger satisfaction for Xgboost model. Online Boarding is the most important feature of this model, which might be crucial for customer experiences and satisfaction. It points toward the efficiency and ease of boarding, which is highly valued by passengers. It is followed by the travel type (personal or business) and by Wi-Fi Service onboard, which shows the importance of both the purpose of travel and access to uninterrupted internet service during flights. The other factors that are believed to be the important contributors to the model include Customer type time or Return, Class-business or economy, and In-flight Entertainment, reflecting the passengers' preference for comfort, service, and entertainment options, respectively. Contributing factors to a minimal influence are Gender, Flight Distance, and Departure Delay shows that these factors do not play an important role in determining overall passenger satisfaction. This hierarchy of feature importance gives actionable insights for airlines on prioritization and optimization tuning of key service areas, hence effective passenger satisfaction.

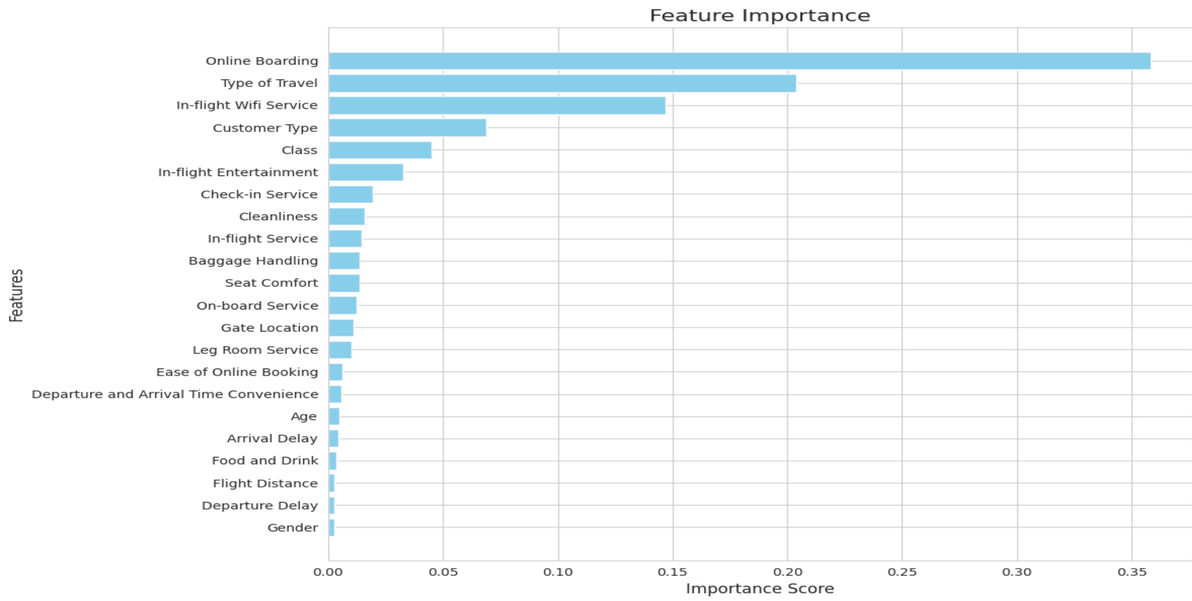

Figure 10: Feature importance graph for XgBoost Model

6.6 Comparison of MCC Values Across Models

In Figure 11, the bar graph represents the MCC values developed in this study in predicting satisfied airline passengers. MCC stands for Matthews correlation coefficient metric that judges model performance taking into consideration all aspects.

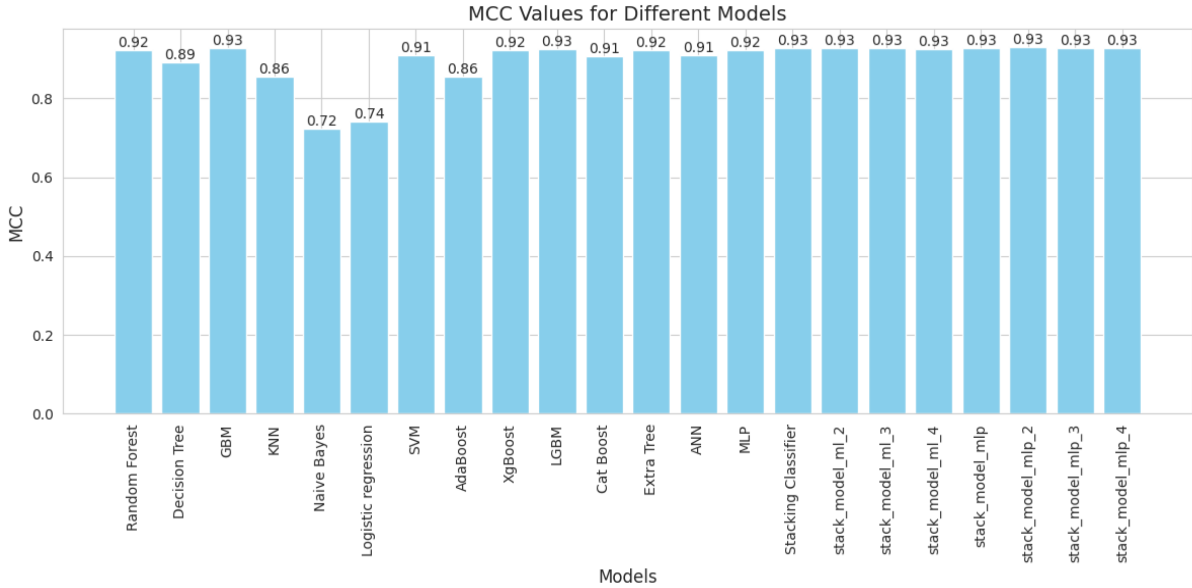

Figure 11: Bar graph showing MCC Values for all models

The Random Forest, SVM, and GBM were able to reach high MCC up to approximately 0.9, unlike the Naive Bayes and K-Nearest Neighbors classifiers which were giving relatively low MCC indicating their poor suitability. All stacking models performed better than all individual models, with stack_model_ml_2 and stack_model_mlp_2 having the highest MCC values. It reflects the power of ensemble learning and also indicates how the MLP-based stacking model is a bit better than the Logistic Regression-based stacking model, which is good in utilizing base model predictions while capturing data more holistically and performing balanced optimal classification.

6.7 Discussion

The results from the experiments indicate that all the ensemble-based stacking models outperform individual models in terms of accuracy, F1 score, and MCC. Among the individual classifiers, XGBoost, Random Forest, and GBM were top performers with an accuracy of 96.17%, 96.23%, and 96.48%, while F1-scores were around 95.60%, 95.65% and 95.93% respectively for all three models. The ensemble methods like boosting and bagging improved the generalization capability and efficiently predicted passenger satisfaction in the airline sector. However, stacking models that combined multiple base learner's predictions with the help of a meta-learner, achieved even better results and showed their potential for improving classification tasks by leveraging the strengths of various algorithms.

The stacking models using Logistic Regression as a meta-learner (stack model ml to stack model ml 4) showed accuracies in the range of 96.39% to 96.50%, with the highest F1-score of 95.97% in stack_model_ml. Also, the stacking models using MLP as a meta-learner (Models stack_model_mlp to stack_model_mlp_4) delivered slightly higher accuracy, with the best-performing model being stack_model_mlp_2 with 96.53% accuracy, 95.98% F1-score, with highest MCC value of 92.99%. This shows that the MLP meta-learner effectively combines base learner predictions like Random Forest, XGBoost, and GBM for capturing nonlinear relationships and thereby enhancing generalization.

From Figure 12, it is clear that stacking models performed well in all metrics, whereas the best performance comes from the stack_model_mlp_2, showing high precision and recall with good reliability.

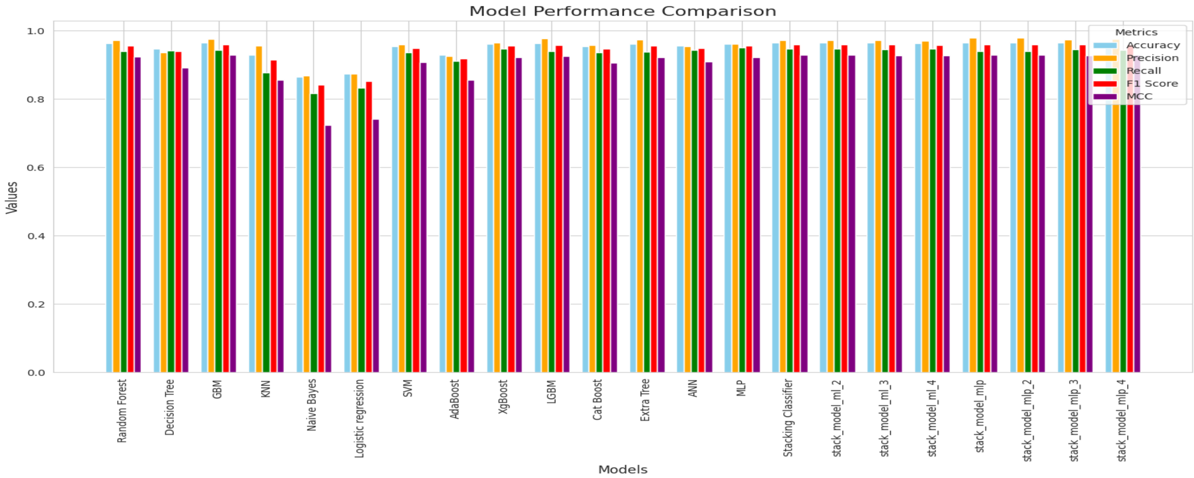

Figure 12: Graph representation of all results

Figure Figure 13 shows the performance of the best-performing model, stack_model_mpl_2, which is represented using the bar graph. It has given immense performance in all key performance indicators. It shows that a model is very robust in doing classification tasks.

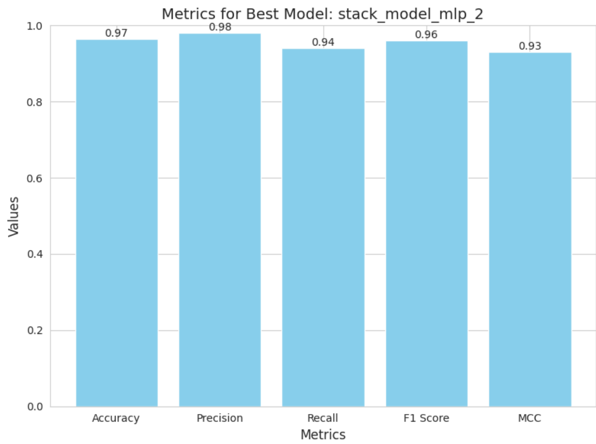

Figure 13: Graph representation of best model

Similarly, Figure 14 radar chart shows a visual representation of how the model performs balanced among these metrics. The shape of the chart is approximately symmetrical, indicating that stack_model_mlp_2 consistently performs very well in all aspects of evaluations, with precision slightly higher and hence making sure it is effective in dealing with the complexities of this dataset.

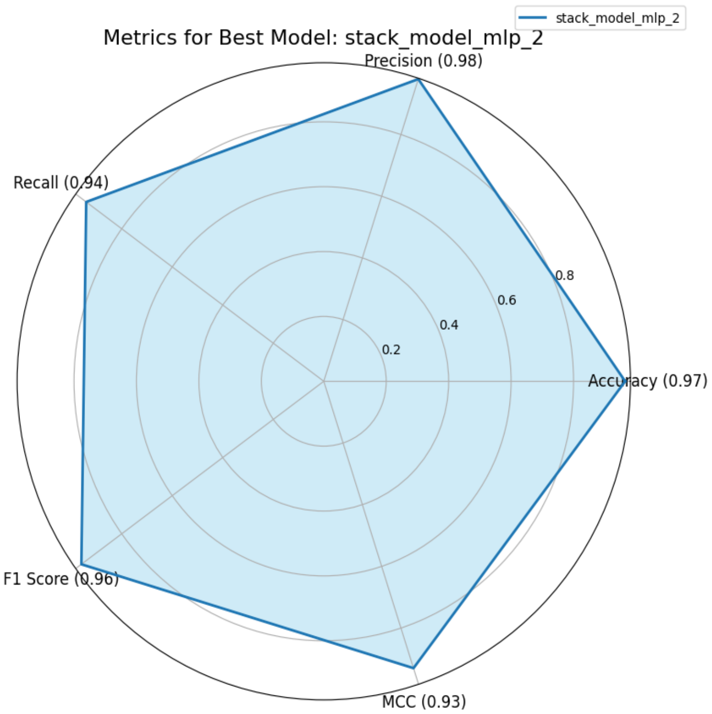

Figure 14: Radar Chart for best model

7 Conclusion and Future Work

This research explored different stacking classifiers and machine learning models for passenger satisfaction prediction in airline sector. The best results were achieved by using stacking classifiers. Accordingly, the best performance reported for the best model - 'stack_model_mlp_2', reached 96.53% accuracy, high precision equal to 97.98%, and a recall of 94.06%, finally reaching up to 95.98% of F1-score, thus ensuring that the model is capable of handling complex data while mining rather for good patterns. Similarly, other models, like XGB and LGBM, performed quite well and showed that most of the time, the ensemble methods yield the best results. However, all these state-of-the-art techniques need a lot of computing power, take up much time for training, and require high-performance hardware to run smoothly.

Key findings in this study identified some important factors that influence passenger satisfaction like onboard Wi-Fi connectivity, streamlining of online boarding processes, and staff behavior towards passengers, especially business or leisure passengers. When such factors improve, it would make flying a great experience for passengers, who thus become loyal, leading the organization to better positioning in the airline industry.

This work may be further directed towards optimizing the stacking frameworks with more advanced hyperparameter tuning and the use of hybrid models, hence allowing for better generalization and performance. Behavioral analytics may also be added, including sentiment analysis, to provide real-time data streams for a better understanding of passenger needs. The airline industry's competition will continue to rise, making the development of robust, scalable, and adaptive predictive frameworks ever-important for sustaining long-term improvement in customer satisfaction.

8 Acknowledgment

Firstly, I would like to sincerely acknowledge my supervisor, Mr. Abdul Qayum, for guiding me during the entire duration of my project and giving me immensely important responses which helped me not only improve my project but also remain motivated towards the completion of my research. A sincere special thanks go to my discussion team members for discussing matters regarding the project with them, thereby enlightening many valuable concepts about it. Thank you, too, NCI library help center-availabilities whatsoever needed by myself in making the project -as well as my dear family and friends who extended consistent motivation whenever necessary.

References

Baydogan, C. and Alatas, B. (2019). Detection of customer satisfaction on unbalanced and multi-class data using machine learning algorithms, *2019 1st international informatics and software engineering conference (UBMYK)*, IEEE, pp. 1–5.

Bellizzi, M. G., Eboli, L., Mazzulla, G. and Postorino, M. N. (2022). Classification trees for analysing highly educated people satisfaction with airlines' services, *Transport Policy* **116**: 199–211.

Dike, S. E., Davis, Z., Abrahams, A., Anjomshoae, A. and Ractham, P. (2024). Evaluation of passengers' expectations and satisfaction in the airline industry: an empirical performance analysis of online reviews, *Benchmarking: An International Journal* **31**(2): 611–639.

Hong, A. C. Y., KHAW, K. W., Chew, X. and Yeong, W. C. (2023). Prediction of us airline passenger satisfaction using machine learning algorithms, *Data Analytics and Applied Mathematics (DAAM)* pp. 7–22.

Hwang, S., Kim, J., Park, E. and Kwon, S. J. (2020). Who will be your next customer: A machine learning approach to customer return visits in airline services, *Journal of Business Research* **121**: 121–126.

Kumar, A., Mehta, V., Rao, T. and Kodipalli, A. (2024). Forecasting airline customer satisfaction through the application of diverse machine learning algorithms, *2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS)*, IEEE, pp. 1–5.

Lee, M. (2024). Comparison of bagging, boosting, and stacking ensemble models for airline customer satisfaction analysis, *FaST-Jurnal Sains dan Teknologi (Journal of Science and Technology)* **8**(1): 76–90.

Mirthipati, T. (2024). Enhancing airline customer satisfaction: A machine learning and causal analysis approach, *arXiv preprint arXiv:2405.09076* .

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

Nurdina, A. and Puspita, A. B. I. (2023). Naive bayes and knn for airline passenger satisfaction classification: Comparative analysis, *Journal of Information System Exploration and Research* **1**(2).

Salah-Ud-Din, M., SS, B. T. L. and Al Ali, H. (2024). Exploratory data analysis and prediction of passenger satisfaction with airline services, *2024 New Trends in Civil Aviation (NTCA)*, IEEE, pp. 295–302.

Shah, F. T., Syed, Z., Imam, A. and Raza, A. (2020). The impact of airline service quality on passengers' behavioral intentions using passenger satisfaction as a mediator, *Journal of Air Transport Management* **85**: 101815.
URL: *https://www.sciencedirect.com/science/article/pii/S0969699717301874*

Song, C., Ma, X., Ardizzone, C. and Zhuang, J. (2024). The adverse impact of flight delays on passenger satisfaction: An innovative prediction model utilizing wide & deep learning, *Journal of Air Transport Management* **114**: 102511.

Suprapto, D. S. and Oetama, R. (2023). Analysis of airline passenger satisfaction using decision tree and naïve bayes algorithms, *Jurnal Informatika Ekonomi Bisnis* pp. 1493–1500.

Wu, S. and Gao, Y. (2024). Machine learning approach to analyze the sentiment of airline passengers' tweets, *Transportation Research Record* **2678**(2): 48–56.