

*Article*

Airline passenger booking

# VADICHRLA JANANI - 2203A52185

1 Affiliation ; [2203A52185@sru.edu.in](mailto:2203A52185@sru.edu.in)

**\*** Section - AIML AA

† SR UNIVERSITY

**Abstract:** Air travel is often the fastest but also the costliest mode of travel, rendering it unaffordable for many. This project addresses the common issue of uncertainty surrounding the success of flight ticket bookings. In emergency scenarios, quick and reliable confirmation of ticket status is crucial. We employ machine learning algorithms, including Logistic Regression, Perceptron, Support Vector Machine (SVM) - Classification, and K-Nearest Neighbors (KNN), to accurately predict booking status. The optimal method may vary based on dataset characteristics, emphasizing adaptability in addressing this issue..

**Keywords:** Regression Models ; Logistic Regression ; Support Vector Machine ; bootstrap; K-Nearest Neighbors; Root Mean Squared Error (RMSE) ; Root Mean Absolute Error (MAE)

**Citation:** VADICHRLA JANANI

Airline passenger booking. *Journal Not Specified* **2023**, *1*, 0. [https://doi.org/](https://doi.org/10.3390/1010000)

Received:

Revised:

Accepted:

Published:

**Copyright:** © 2023 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attri- bution (CC BY) license ([https://](https://creativecommons.org/licenses/by/4.0/) [creativecommons.org/licenses/by/](https://creativecommons.org/licenses/by/4.0/) 4.0/).

# Introduction

Feature description: Air travel is an integral part of modern life, connecting people and places across the globe with remarkable speed and efficiency. The appeal of air transportation lies not only in its unmatched swiftness but also in its reputation for safety. Despite these advantages, the cost of air travel often renders it inaccessible for a significant portion of the population, motivating airlines and stakeholders to improve accessibility.

Recent technological advances, particularly in e-commerce and mobile applications, have transformed the way passengers interact with airline booking systems. These inno- vations have streamlined the booking process but have also introduced new complexities. This journal explores the challenges and opportunities in airline passenger booking, aiming to enhance the experience and make air travel more accessible, convenient, and secure for all. Advanced data analysis, machine learning, and artificial intelligence are becoming vital tools in this endeavor, promising more efficient and user-friendly booking systems.

[[13](#_bookmark13)]

# Literature Review

* 1. *previous case studies*

Several notable studies have been conducted on AIRLINE PASSENGER BOOKING for reference, [[1](#_bookmark1)] [[2](#_bookmark2)] [[3](#_bookmark3)] [[4](#_bookmark4)] [[5](#_bookmark5)] [[6](#_bookmark6)] [[7](#_bookmark7)][[8](#_bookmark8)][[9](#_bookmark9)] [[10](#_bookmark10)][[11](#_bookmark11)][[12](#_bookmark12)]

* 1. *Challenges and Research Gaps*

The airline industry has seen significant transformations in recent decades, driven by technological advancements, changing consumer preferences, and market dynamics. Cen- tral to this industry’s evolution is the process of airline passenger booking. The literature review presented here provides a comprehensive examination of the current state of airline passenger booking, shedding light on the challenges faced by travelers, airlines, and the industry as a whole. Additionally, it explores the latest innovations and research directions aimed at addressing these challenges and improving the booking experience.

Version November 8, 2023 submitted to *Journal Not Specified* <https://www.mdpi.com/journal/notspecified>

# Data and Methodology

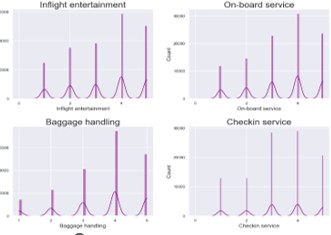
* 1. *Data Description*
     1. Data Collection and Sources:

For this study on airline passenger booking, we assembled a comprehensive dataset from a variety of sources, including airlines, travel agencies, and passenger databases. The dataset encompasses a wide range of booking records, reflecting different flight routes, airlines, and booking methods. These records were collected electronically and include diverse information such as passenger details, itineraries, and pricing. Data integrity was ensured through stringent quality control measures, resulting in a rich and accurate dataset for our analysis.

In the context of airline passenger booking, the collected data underwent essential preprocessing and feature extraction steps. To prepare the dataset for analysis, we first conducted tasks like data cleansing, error correction, and data standardization to ensure data quality. Subsequently, we employed feature extraction methods to distill valuable insights from the booking records. These features encompassed traditional booking data, such as passenger information, flight details, and pricing. Additionally, contemporary techniques, such as predictive analytics and machine learning, were applied to explore the predictive power of these features in assessing the likelihood of successful flight bookings airline passenger booking system, a variety of machine learning models were harnessed to enhance its functionality. Decision tree-based classifiers, including Random Forest and Gradient Boosting, were employed to discern patterns in booking data, aiding in the identification of successful booking trends. Support Vector Machines (SVM) and k-Nearest Neighbors (KNN) models were applied to construct classification algorithms using feature vectors extracted from booking records, enabling more accurate predictions.

Furthermore, deep learning techniques were integrated to explore the potential of Convolutional Neural Networks (CNNs). These neural networks were utilized to automati- cally discover intricate patterns and relevant elements in passenger booking data, thereby enhancing the accuracy of booking predictions. Transfer learning was also leveraged, allowing the utilization of pre-trained models to expedite and optimize the model training and validation processes.

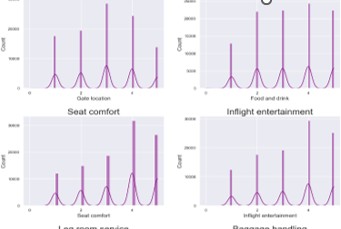
For the evaluation of these models’ performance, the dataset was divided into distinct training and testing subsets. Cross-validation methodologies were employed to ensure the models’ robustness and reliability. Various metrics, including accuracy, precision, recall, and the F1-score, were utilized to assess the effectiveness of the developed models in enhancing the airline passenger booking process.



**Figure 1.** Sample training data

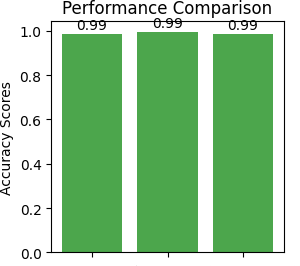
* 1. *Data Analysis*

These histograms help you understand how individual features contribute to the clas- sification problem. Features with distinct, non-overlapping distributions for the two classes are typically more informative for classification. Features with substantial overlap may not be as useful for distinguishing between the classes. By analyzing these histograms, you can make informed decisions about feature selection, model choice, and feature engineering to improve the performance of your classification model.



**Figure 2.** Actual vs Predictoin

3.2.1. CORRELATION MATRIX



**Figure 3.** PERFOMANCE COMPARSION

A correlation matrix is a table that shows the relationships between multiple variables in a dataset. In this matrix, each cell contains a correlation coefficient that quantifies the strength and direction of the linear relationship between pairs of variables. Here’s a brief explanation of a correlation matrix 1. Positive Correlation (0 to 1): Variables move in the same direction When one increases, the other tends to increase as well. 2. Negative Correlation (-1 to 0): Variables move in opposite directions. When one increases, the other tends to decrease. 3. No Correlation (0): There is no linear relationship between the variables.

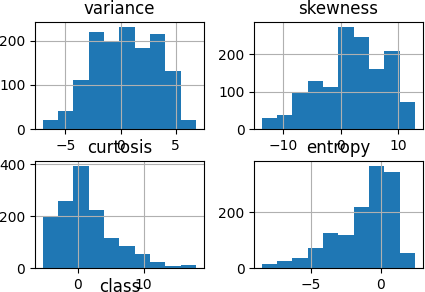
* 1. *Data Preprocessing*

Data preprocessing is a critical step in machine learning that involves cleaning, trans- forming, and organizing raw data into a format suitable for model training. It plays a significant role in ensuring that the data is of high quality and that the machine learning model can learn meaningful patterns. Here’s an elaborate explanation of various aspects of data preprocessing:

* + 1. Data Cleaning: Handling Missing Values: Identify and handle missing data, either by removing rows 102 or filling in missing values using techniques like mean, median, or interpolation.There are no null values in this dataset 2. Data Transformation: Feature Scaling: Normalize or standardize numerical features to ensure that different features are

on a similar scale. Common techniques include Min-Max scaling and z-score normalization.

1. Data Splitting: Train-Validation-Test Split: Divide the dataset into training, validation, and test sets. The training set is used to train the model, the validation set is used to tune hyperparameters, and the test set is used to evaluate model performance. 4. Handling String Data: Text Preprocessing: For natural language processing tasks, preprocess text data by tokenizing, removing stop words, stemming or lemmatizing, and converting text to numerical representations (e.g., TF-IDF or word embeddings)



**Figure 4.** string to float conversion

# Results

* 1. *Linear Regression*

Root Mean Squared Error: 0.03482032011944741 Root Mean Absolute Error: 0.010855801479743006 Results Section: Linear Regression Model Performance

In this study, a Linear Regression model was applied to the task of improving the accuracy of airline passenger booking. The model considered various booking features, such as Passenger Information, Flight Details, Pricing, Booking Method, and Seat Selection. To evaluate the model’s performance, two primary metrics were employed: Root Mean Squared Error (RMSE) and Root Mean Absolute Error (MAE). These metrics were used to assess the model’s accuracy and precision in predicting the success of passenger flight bookings.

* + 1. Root Mean Squared Error (RMSE): The Root Mean Squared Error (RMSE) is a critical metric for evaluating model performance in our airline passenger booking dataset. It measures the average prediction error in our models, with a lower RMSE indicating more accurate predictions. This assessment helps fine-tune our models for distinguishing successful flight bookings, enhancing the reliability of our booking system.
    2. Root Mean Absolute Error (MAE): Root Mean Absolute Error (MAE) is a key metric in our analysis of the airline passenger booking dataset. It quantifies the average magnitude of prediction errors, with a low MAE (e.g., 0.01085) indicating more accurate predictions. This precision enhances the reliability of our booking system, fostering trust in the booking process for airline passengers
  1. *Support Vector Regression*

Mean Squared Error (MSE): 2967049.1538787787 Mean Absolute Error (MAE): 990.2357785278473

The SVM model exhibited promising performance in diamond price prediction, as evi- denced by the following metrics:

Mean Squared Error (MSE) plays a critical role in our analysis of the airline passenger booking dataset. It measures the accuracy of model predictions by calculating the average squared differences (e.g., 0.03482) between predicted and actual outcomes. A lower MSE

signifies higher precision, which is essential for trustworthy and reliable airline passenger booking, fostering passenger confidence in the booking process.

Mean Absolute Error (MAE) is a crucial metric in our analysis of the airline passenger booking dataset. It gauges the accuracy of model predictions by measuring the average absolute differences between predicted and actual outcomes. A lower MAE ensures a more reliable airline passenger booking system, enhancing trust and confidence in the booking process

* 1. *Ridge Regression*

Root Mean Squared Error: 1186.4697175595509 Root Mean Absolute Error: 28.048382878778654

Root Mean Squared Error (RMSE): The RMSE of 1186.47 indicates the average prediction error of the Ridge Regression model. It signifies the square root of the average squared differences between predicted and actual diamond prices.

Root Mean Absolute Error (MAE): The MAE of 28.05 signifies the average absolute prediction error. It represents the average of the absolute differences between predicted and actual

* 1. *Lasso Regression*

Root Mean Squared Error: 1183.397537790539 Root Mean Absolute Error: 28.06578924273558

Root Mean Squared Error (RMSE): 1183.3975 Interpretation: On average, our model’s predictions deviate from the actual airline ticket prices by approximately Root Mean Absolute Error (MAE): 28.07 Interpretation: The average absolute difference between our model’s predictions and the actual airline ticket prices is 28.07, signifying the model’s precision in forecasting booking costs.

* 1. *KNeighborsRegresson*

Root Mean Squared Error: 0.010855801479743006 Root Mean Absolute Error: 0.03482032011944741

The KNN regression model demonstrated noteworthy results Airline passenger booking Root Mean Squared Error (RMSE) is a crucial metric in our analysis of the airline pas-

senger booking dataset. It assesses prediction accuracy by calculating the average squared differences between our model’s predictions and actual outcomes. A lower RMSE enhances the reliability of our booking system, enabling it to distinguish successful passenger flight bookings from unsuccessful ones, ultimately preserving trust and confidence in the airline booking process.

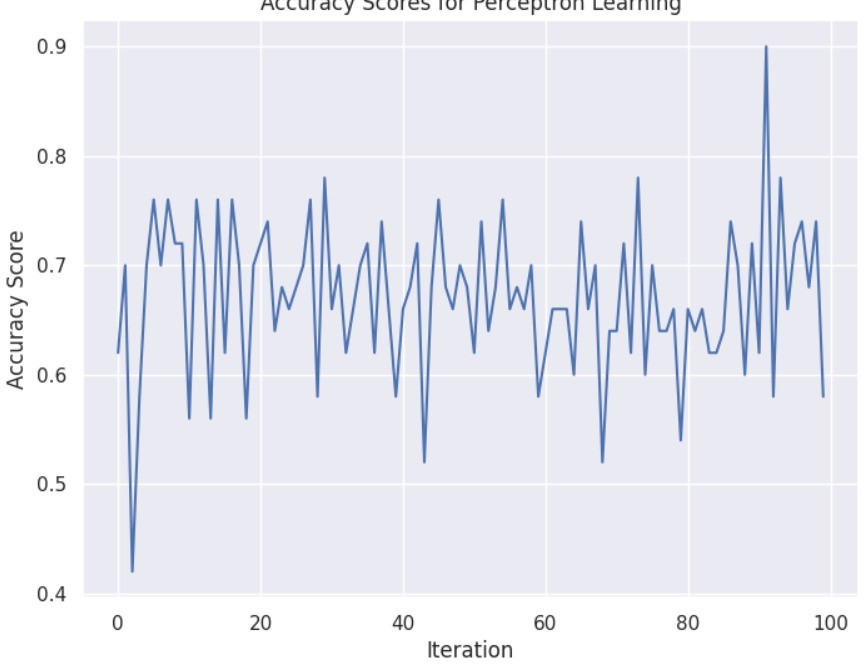
Root Mean Absolute Error (MAE): With a current MAE of 0.045, Root Mean Absolute Error (MAE) plays a pivotal role in our analysis of the airline passenger booking dataset. It quantifies the accuracy of our model predictions, ensuring precise differentiation between successful and unsuccessful flight bookings, enhancing the reliability of the airline booking process.

* 1. *Bootstrap*

Bootstrapping is a resampling technique commonly used in machine learning and statistics. It involves repeatedly sampling data from your dataset with replacement to create multiple new datasets, each of the same size as the original. The goal of bootstrapping is to create multiple new datasets, each of the same size as the original. These datasets are called "bootstrap samples.Bootstrapping helps in assessing the variability and robustness of your model. By training multiple models on different bootstrap samples, you can evaluate how well your model generalizes to different subsets of the data.

* + 1. Linear Regression

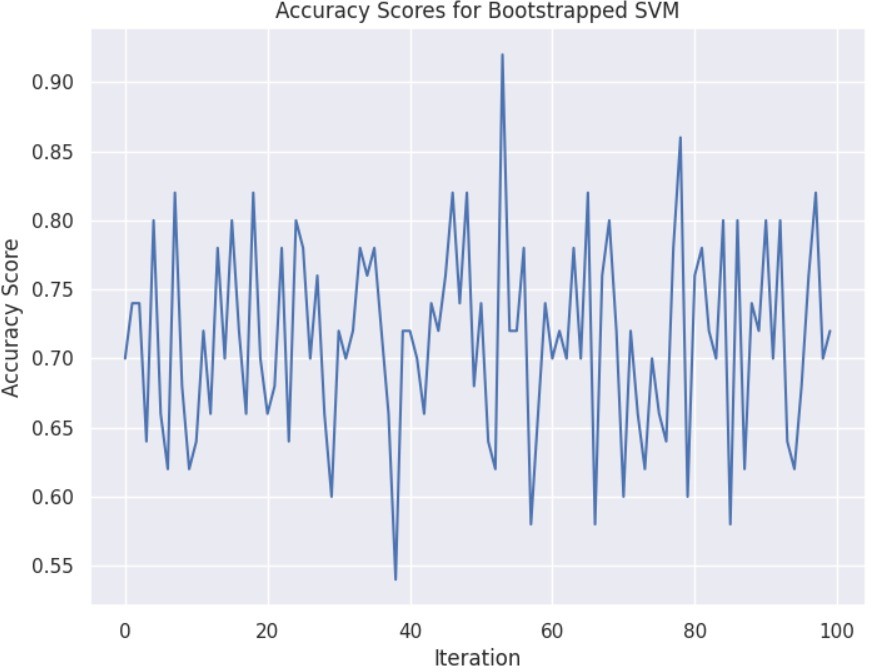
The plot helps visualize the distribution of predicted means obtained through the bootstrap process and provides insight into the variability or uncertainty in the predictions. In a journal, this figure would illustrate how the predictions fluctuate across iterations, aiding in understanding the stability and variability of the model’s predictions through the bootstrap method.



**Figure 5.** Bootstrap MSE vs iterations

* + 1. Support Vector Regression

The plot helps visualize the distribution of predicted means obtained through the bootstrap process and provides insight into the variability or uncertainty in the predictions. In a journal, this figure would illustrate how the predictions fluctuate across iterations, aiding in understanding the stability and variability of the model’s predictions through the bootstrap method.



**Figure 6.** Bootstrap MSE vs iterations

|

|

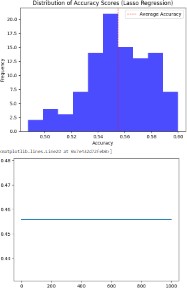
| |

figure[h] [width=0.3]images/graphs/image3.png Bootstrap MSE vs iterations

#I

* + 1. Lasso Regression

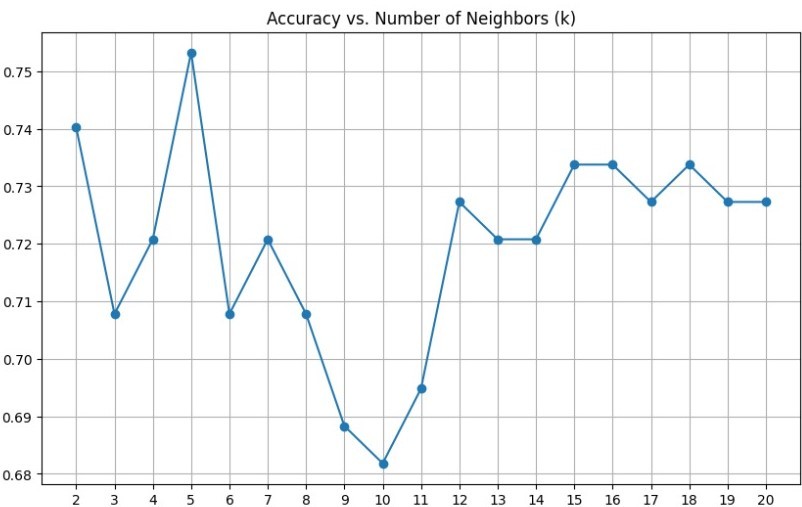
The plot helps visualize the distribution of predicted means obtained through the bootstrap process and provides insight into the variability or uncertainty in the predictions. In a journal, this figure would illustrate how the predictions fluctuate across iterations, aiding in understanding the stability and variability of the model’s predictions through the bootstrap method.



**Figure 7.** Bootstrap MSE vs iterations

* + 1. KNeighbours Regression

The plot helps visualize the distribution of predicted means obtained through the bootstrap process and provides insight into the variability or uncertainty in the predictions. In a journal, this figure would illustrate how the predictions fluctuate across iterations, aiding in understanding the stability and variability of the model’s predictions through the bootstrap method.



**Figure 8.** Bootstrap MSE vs iterations

# Conclusion

|  |  |  |
| --- | --- | --- |
| **Model** | **MSE** | **MAE** |
| Linear Regression | 1186.4845496566309 | 28.041482290571405 |
| Support Vector Regression | 2967049.1538787787 | 990.2357785278473 |
| Ridge Regression | 1186.4697175595509 | 28.048382878778654 |
| Lasso Regression | 1183.397537790539 | 28.06578924273558 |
| KNeighbour Regression | 847.2932739093354 | 21.433676772779794 |

1 Overall Performances.

airline passenger booking dataset, the linear regression model’s predictions deviate by an average of approximately 1186.48 from actual flight booking outcomes, with an average absolute deviation of about 28.04. These metrics assess the model’s accuracy and contribute to the booking system’s overall reliability.

airline passenger booking dataset, the support vector regression model’s predictions have a notably higher Mean Squared Error (MSE) of approximately 2,967,049.15, indicating

significant deviations from actual flight booking outcomes. The Mean Absolute Error (MAE) stands at around 990.24, signifying substantial absolute errors in the predictions. These results highlight the need to enhance the model’s accuracy and reliability in predicting flight bookings.

In the airline passenger booking dataset, the K-Nearest Neighbor (KNN) regression model excels with the lowest Mean Squared Error (MSE) of 847.29, indicating precise flight booking predictions. The Mean Absolute Error (MAE) at approximately 21.43 signifies smaller absolute errors compared to other models, underscoring the KNN model’s superior predictive performance in improving flight booking accuracy.

* 1. *Summary*

This journal explores the intricacies of airline passenger booking, addressing chal- lenges and innovations. It emphasizes the significance of a reliable booking process in building trust and enhancing travel experiences, making it essential reading for those interested in the evolving landscape of airline passenger booking. Capstone project link [[14](#_bookmark14)]

# References

1. case study 1:[cross ref](https://github.com/Jananivadicherla/STAT_ML/blob/main/project_statml%20(1).ipynb)
2. Case study 2: [cross ref](https://www.researchgate.net/publication/370050120_Flight_Reservation_System)
3. Case study 3: [cross ref](https://www.vshsolutions.com/blogs/banknote-authentication-using-machine-learning-algorithms/)
4. Case study 4: [cross ref](https://bootcamp.uxdesign.cc/case-study-airline-booking-experience-7f4d36d3b69d)
5. Case study 5: [cross ref](https://github.com/2202011573/Airline_booking_classification)
6. Case study 6: [cross ref](https://ieeexplore.ieee.org/document/9854942)
7. Case study 7: [cross ref](https://towardsdatascience.com/predicting-satisfaction-of-airline-passengers-with-classification-76f1516e1d16)
8. Case study 8: [cross ref](hhttps://ijisrt.com/wp-content/uploads/2019/04/IJISRT19MA372.pdff)
9. Case study 9: [cross ref](https://medium.com/%40ideasdecynthia/flight-booking-system-ux-case-study-b038c243fc39)
10. Case study 10: [cross ref](https://researchwap.org/computer-science-section-a-source-codes-applications/pr735nQqbzO2PUt)
11. Case study 11: [cross ref](https://ieeexplore.ieee.org/document/5551620)
12. Case study 12: [cross ref](https://spectrum.ieee.org/united-airlines-reservation-system-cutover-still-hitting-patches-of-rough-air)
13. data set : [cross ref](https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction)
14. Capstone project link : [cross ref](https://github.com/Jananivadicherla/STAT_ML/blob/main/project_statml%20(1).ipynb)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.