Airline Passenger Satisfaction Analysis

MACHINE LEARNING & POWER BI DASHBOARD PROJECT

PROJECT REPORT

BY

AKHILA KS

DATA SCIENCE

UNDER THE GUIDANCE & SUPERVISION OF

AMRITHA MR

Palakkad,kerala,678702

CONTENTS

1.INTRODUCTION

The airline industry is one of the most service-oriented and competitive sectors in the global economy. Customer satisfaction directly influences brand loyalty, repeat purchases, and overall profitability of airline companies. With increasing passenger expectations and growing competition, airlines must continuously monitor and improve service quality. Traditional analysis methods are often insufficient to handle large volumes of passenger data. Therefore, machine learning techniques provide an efficient and accurate way to analyze customer behavior and predict satisfaction levels.

This project focuses on applying machine learning algorithms to predict airline passenger satisfaction based on demographic details, travel information, and service quality ratings. The insights generated from this analysis can help airline management make data-driven decisions to improve customer experience.

2.ABSTRACT

This project presents a comprehensive machine learning approach to predict airline passenger satisfaction. The Airline Passenger Satisfaction dataset was used to perform data preprocessing, exploratory data analysis (EDA), feature engineering, and model development. Several classification algorithms including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naive Bayes were implemented and evaluated.

The models were assessed using accuracy and confusion matrix metrics. Among all the models, the Random Forest classifier achieved the highest accuracy of 96.32%, indicating superior predictive performance. An interactive dashboard was also developed using data visualization tools to present insights in a user-friendly manner. The results demonstrate the effectiveness of machine learning in predicting customer satisfaction and identifying key factors influencing passenger experience.

3.OBJECTIVES

The primary objectives of this project are:

- To understand airline passenger behavior through data analysis
- To preprocess and clean real-world airline passenger data
- To perform exploratory data analysis and identify important trends
- To build and compare multiple machine learning classification models
- To evaluate models using performance metrics such as accuracy
- To select the best-performing model for prediction
- To visualize insights using an interactive dashboard
- To enable prediction of passenger satisfaction based on user input

4.METHODOLOGY

The project follows a structured machine learning workflow consisting of the following steps:

1. Data Collection

The Airline Passenger Satisfaction dataset was collected from a publicly available source. The dataset contains detailed information about passengers, including demographic details, travel-related attributes, service quality ratings, and satisfaction labels.

2. Data Cleaning

- Missing values in the "Arrival Delay in Minutes" column were handled using median imputation.
- Duplicate records were identified and removed to avoid bias in model training.
- Data consistency was ensured before further processing.

3. Data Encoding

Categorical variables such as Gender, Customer Type, Type of Travel, Class, and Satisfaction were converted into numerical values using Label Encoding, enabling machine learning algorithms to process them effectively.

4. Feature Scaling

Numerical features were standardized using Standard Scaler to bring all features to a common scale. This step is particularly important for distance-based and margin-based algorithms such as KNN and SVM.

5. Train-Test Split

The dataset was split into training and testing sets using an 80:20 ratio. The training data was used to build models, while the testing data was used to evaluate model performance.

6. Model Development

Multiple machine learning classification models were implemented, trained, and compared.

7. Model Evaluation

Each model was evaluated using accuracy score and confusion matrix to measure classification performance.

5.DATASET OVERVIEW

- Dataset Name: Airline Passenger Satisfaction Dataset
- Total Records: Approximately 130,000
- Target Variable: Satisfaction (Satisfied / Neutral or Dissatisfied)

- Feature Categories:

    - Demographic Features: Age, Gender
    - Travel Features: Class, Type of Travel, Flight Distance
    - Service Quality Ratings: Seat Comfort, Inflight Entertainment, Cleanliness, Online Boarding
    - Delay Information: Departure Delay, Arrival Delay

The dataset provides a rich source of information to analyze factors affecting passenger satisfaction.

6.DATA ANALYSIS & EDA VISUALS

Exploratory Data Analysis (EDA) was conducted to understand the structure and characteristics of the dataset.

EDA Techniques Used:

- Descriptive statistics to summarize numerical features
- Visualization of satisfaction distribution
- Analysis of satisfaction across travel classes and customer types
- Correlation analysis between service ratings and satisfaction
- Delay impact analysis

Key Insights:

- Business class passengers exhibited higher satisfaction compared to economy class
- Loyal customers were more likely to be satisfied
- Service quality ratings had a strong positive correlation with satisfaction
- Increased delays negatively affected passenger satisfaction

Visualizations such as bar charts, pie charts, and heatmaps were used to support these findings.

7.MODEL DEVELOPMENT & RESULTS

The following machine learning algorithms were developed and evaluated:

1. Logistic Regression
2. Decision Tree Classifier
3. Random Forest Classifier
4. Support Vector Machine (SVM)
5. K-Nearest Neighbors (KNN)
6. Naive Bayes

Model Performance Comparison:

- Logistic Regression Accuracy: ~87%
- Decision Tree Accuracy: ~94%
- Random Forest Accuracy: 96.32%
- SVM Accuracy: ~90%
- KNN Accuracy: ~88%
- Naive Bayes Accuracy: ~86%

Random Forest outperformed all other models due to its ensemble nature and ability to handle complex relationships. A confusion matrix was used to further validate model predictions.

8.DASHBOARD OVERVIEW

An interactive dashboard was designed using Power BI / Tableau / Excel to present insights visually.

Dashboard Features:

- KPI cards displaying total passengers, satisfied passengers, and satisfaction percentage
- Pie chart showing satisfaction distribution
- Bar charts showing satisfaction by class and type of travel
- Heatmap visualizing the impact of service quality ratings
- Interactive filters for gender, class, and customer type

The dashboard enables stakeholders to explore data dynamically and supports informed decision-making.

9.CONCLUSION & FUTURE SCOPE

Conclusion

This project successfully applied machine learning techniques to predict airline passenger satisfaction. The Random Forest model achieved the highest accuracy of 96.32%, making it the most suitable model for this task. The analysis revealed that service quality factors play a critical role in determining passenger satisfaction.

Future Scope

- Implementation of deep learning models for enhanced prediction
- Real-time satisfaction prediction using live airline data
- Deployment of the model using Streamlit or Flask
- Integration of customer feedback and sentiment analysis
- Expansion of dashboard capabilities with real-time updates