Airline Passenger Satisfaction Analysis - Abstract

Abstract: Airline Passenger Satisfaction Analysis

Passenger satisfaction is a critical metric in the aviation industry, directly influencing customer loyalty, airline reputation, and profitability. The "Airline Passenger Satisfaction Analysis" project aims to analyze the primary factors that contribute to passenger satisfaction using a structured data-driven approach. Leveraging a publicly available airline passenger dataset, the project utilizes data preprocessing, exploratory data analysis, and machine learning models to uncover the variables that most significantly impact satisfaction and predict passenger sentiments effectively.

Problem Statement and Overview

The aviation industry collects vast amounts of passenger data, including travel class, flight distance, age, in-flight services, and customer feedback. However, the challenge lies in extracting actionable insights to understand what truly drives passenger satisfaction. The project focuses on addressing this gap by analyzing these factors systematically and building a predictive model to classify passenger satisfaction as "satisfied" or "neutral/dissatisfied." The ultimate objective is to provide airlines with insights that can enhance service quality and operational decision-making.

Tools and Applications Used

The analysis was performed using Python within a Jupyter Notebook environment, facilitating an iterative workflow and clear documentation of outputs and visualizations. Key libraries and tools used include:

- Pandas: For data cleaning, preprocessing, and handling categorical data.
- NumPy: For numerical operations.
- Matplotlib and Seaborn: For generating clear, insightful visualizations.
- Scikit-learn: For machine learning model building, training, and evaluation.
- LabelEncoder: For transforming categorical labels into numerical values.
- StandardScaler: For feature scaling to normalize data before model training.

Detailed Description of Sub Modules

1. Data Preprocessing: Null values were checked, and categorical columns such as "Gender" and "Customer Type" were encoded using LabelEncoder. The "id" column was dropped as it was not relevant for modeling.

2. Exploratory Data Analysis (EDA): Distributions of satisfied vs. dissatisfied passengers were visualized, and the relationships between variables such as travel type, class, and age with satisfaction were analyzed using bar plots and histograms. Correlation matrices helped identify the strength of relationships between numerical features and satisfaction.

3. Feature Scaling: StandardScaler was applied to numerical features to bring them to a standard range, improving model stability and convergence during training.

4. Model Building: A Random Forest Classifier was used for classification, chosen for its ability to handle mixed feature types and provide insights into feature importance. The dataset was split into training and test sets (typically in an 80-20 split) to evaluate model performance.

5. Model Evaluation: Accuracy and classification reports were generated, measuring precision, recall, and F1-score to assess the effectiveness of the model in predicting passenger satisfaction. The Random Forest Classifier achieved high accuracy, indicating good model performance.

Design or Flow of the Project

The project followed this structured pipeline:

1. Problem Definition 2. Data Loading 3. Data Cleaning & Preprocessing 4. EDA & Visualization 5. Feature Engineering & Scaling 6. Model Building (Random Forest) 7. Evaluation & Validation 8. Insight Generation & Reporting.

This workflow ensured systematic progression from raw data to insight generation while maintaining interpretability.

Conclusion or Expected Output

The project successfully identified key drivers of airline passenger satisfaction, including travel class, type of travel, age, and service ratings (e.g., in-flight service, cleanliness). The Random Forest model accurately predicted passenger satisfaction and highlighted which variables were most impactful, providing actionable insights for airlines to enhance customer experience. Airlines can use these findings to prioritize improvements, such as enhancing in-flight services or focusing on premium passenger experience, thereby increasing customer loyalty and operational efficiency.

Overall, the "Airline Passenger Satisfaction Analysis" project demonstrates how data analytics and machine learning can transform raw passenger data into actionable strategies for improving satisfaction and business outcomes within the airline industry.