

Airline Passenger Satisfaction Analysis

Overview

This project involves analyzing airline passenger satisfaction using machine learning techniques. The primary objective is to predict passenger satisfaction based on various factors influencing their flight experience. The analysis includes several key stages:

1. **Data Cleaning and Preprocessing:** Involves handling missing values, removing irrelevant columns, and formatting the data to ensure it is suitable for analysis and modeling.
2. **Exploratory Data Analysis (EDA):** Utilizes data visualization to uncover trends and patterns. Key areas of focus include the relationship between passenger satisfaction and factors such as type of travel, class, and in-flight services.
3. **Statistical Analysis:** Employs hypothesis tests to validate observations made during the EDA phase. This includes assessing the statistical significance of factors like travel type, class, and gender in influencing satisfaction levels.
4. **Feature Engineering:** Focuses on encoding categorical variables to transform them into a format that machine learning models can effectively process.
5. **Model Training and Evaluation:** Involves training and evaluating the performance of various machine learning models to predict passenger satisfaction. Models utilized include Random Forest, XGBoost, and K-Nearest Neighbors (KNN).

Key Findings

- Business travelers exhibit higher satisfaction levels compared to personal travelers.
- Passengers traveling in higher classes tend to report greater satisfaction.
- The quality of in-flight services significantly correlates with passenger satisfaction.
- Delays in departure and arrival have a negative impact on overall satisfaction.

Models Used

- Random Forest: Achieved an accuracy of 96.28%.
- XGBoost: Demonstrated an accuracy of 96.17%.
- K-Nearest Neighbors (KNN): Obtained an accuracy of 92.99% with k=7.

Tools and Libraries Used

- Python: Primary programming language.
- Pandas: For data manipulation and analysis.
- Matplotlib and Seaborn: For data visualization.
- Scikit-learn: For machine learning.
- SciPy: For statistical analysis.
- XGBoost: For gradient boosting.

Dataset Description

The dataset includes the following columns:

- 'Gender': Gender of the passenger.
- 'Customer Type': Type of customer (e.g., returning or first-time).
- 'Age': Age of the passenger.
- 'Type of Travel': Purpose of travel (e.g., business or personal).
- 'Class': Travel class (e.g., Economy, Business).
- 'Flight Distance': Distance of the flight.
- 'Inflight wifi service': Rating of inflight wifi service.
- 'Departure/Arrival time convenient': Rating of departure/arrival time convenience.
- 'Ease of Online booking': Rating of online booking ease.
- 'Gate location': Rating of gate location.
- 'Food and drink': Rating of food and drink.
- 'Online boarding': Rating of online boarding.
- 'Seat comfort': Rating of seat comfort.
- 'Inflight entertainment': Rating of inflight entertainment.
- 'On-board service': Rating of on-board service.
- 'Leg room service': Rating of leg room service.
- 'Baggage handling': Rating of baggage handling.
- 'Checkin service': Rating of check-in service.
- 'Inflight service': Rating of inflight service.
- 'Cleanliness': Rating of cleanliness.
- 'Departure Delay in Minutes': Departure delay in minutes.
- 'Arrival Delay in Minutes': Arrival delay in minutes.
- 'Satisfaction': Passenger satisfaction level (satisfied or dissatisfied).

How to Run the Project

1. Ensure you have Python installed (preferably Python 3.6 or higher).
2. Install the required libraries using:
3. `pip install pandas numpy matplotlib seaborn scikit-learn scipy xgboost`
4. Open and run the Jupyter Notebook (de-airlines-v1-1-2.pdf) to reproduce the analysis.

Conclusion

This project provides a comprehensive analysis of airline passenger satisfaction, identifying key factors that influence it. The insights derived can be used to inform strategies for improving passenger experience and overall satisfaction.