

Big Data Project Proposal

Team #12

Airline Passenger Satisfaction Prediction

Team members:

Name	Section	BN
Donia Gameel	1	24
Shaza Mohammed	1	32
Heba Ashraf Raslan	2	32

Idea & Motivation:

The project aims to predict airline passenger satisfaction based on various factors such as flight distance, in-flight service, ease of online booking, and departure/arrival time convenience.

Understanding the factors that contribute to passenger satisfaction is crucial for airlines to improve their services, enhance customer experience, and increase customer loyalty. By analyzing this dataset, we aim to provide valuable insights into what drives passenger satisfaction and how airlines can better meet customer expectations.

Dataset: [[Link](#)]

The dataset contains information about airline passengers' demographics, travel preferences, and satisfaction ratings. It includes features such as age, gender, travel class, inflight Wi-Fi service, ease of online booking, departure/arrival time convenience, and passenger satisfaction rating.

Planned Approach:

1. Data Preprocessing: Clean the dataset, handle missing values, and encode categorical variables.
2. Exploratory Data Analysis (EDA): Conduct EDA to understand the distribution of variables, identify patterns, and explore relationships between variables.

We will visualize the distribution of passenger satisfaction ratings and explore how different features such as flight distance, inflight service, and departure/arrival time convenience relate to passenger satisfaction. We will also analyze the correlation between features to identify potential predictors of satisfaction.

3. Feature Selection: Select relevant features for the prediction model based on EDA findings and domain knowledge.
4. Model Selection: Evaluate different machine learning models for classification, such as Logistic Regression, Random Forest, Decision Tree, and Naive Bayes using **Map-Reducer Spark**.
5. Model Training: Train the selected model on the dataset, using a portion of the data for training and another portion for validation.
6. Model Evaluation: Evaluate the trained model using metrics such as accuracy, precision, recall, and F1-score.
7. Prediction: Use the trained model to predict passenger satisfaction for new data.

Algorithm with MapReduce:

For the MapReduce implementation, we will use the Naive Bayes algorithm. We may explore other algorithms if needed, based on the performance.