EAS 508: Project Preliminary Report

Group 31: Sai Teja Dondeti, Sai Teja Sankarneni, Nishant Varma Indukuri, Subba Rao Koduri

Topic: Flight Fare Prediction

Introduction:

The objective of the study is to analyze the flight fare prediction dataset and to conduct various statistical hypothesis tests in order to get meaningful information from it. With this dataset, we will predict the flight ticket price. The price may rely upon different factors. Each factor has its own proprietary rules and algorithms to set the price accordingly. We will use the advances in Machine Learning (ML) to infer such rules and model the price variation.

Data Description:

Data was collected in two parts: one for economy class tickets and another for business class tickets. A total of 300261 distinct flight booking options were extracted. Data was collected for 50 days, from February 11th to March 31st, 2022 for flight travel between India's top 6 metro cities.

Proposed Analysis:

Firstly, we will analyze individually all the factors such as type of airline, class, number of stops etc., that influence the flight fare. Then we will model the data using several models and obtain the best performing model and train the model to predict the flight fares.

Analysis Methods that will be used:

We will use several models like Linear Regression, Ridge Regression, Lasso regression, Bagging Regressor, Random Forest Regressor, Extra Trees Regressor and obtain the best model for predicting the flight fare based on factors like MSE, R Square, Adjusted R Square etc.

Milestones:

Analyzing the data thoroughly will aid in the discovery of factors that really affects the flight fare and will provide knowledge to the passengers regarding the factors to keep in mind for booking flight tickets and obtain the best model for predicting flight fares.

References:

Data Set: Flight Price Prediction | Kaggle

Regression in Python: Linear Regression in Python – Real Python