PREDICT FLIGHT DELAYS USING SUPERVISED MACHINE LEARNING TECHNIQUE

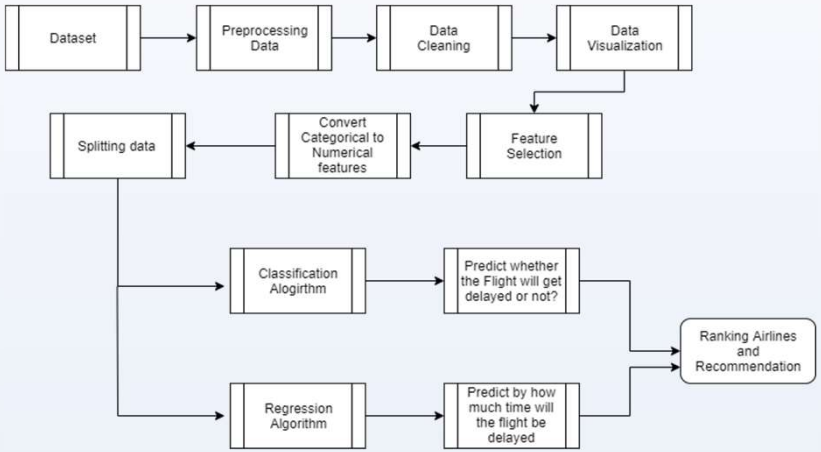
ABSTRACT

- In this project, we analyzed the various factors responsible for flight delays and applied machine learning models to predict whether a given flight would be delayed or not. Also with certain features we can predict how far the delay is going to be using some regression techniques like Random Forest Regression and Decision Tree Regression.
- We also added a recommendation feature in which given a source and destination, we would list flights which are recommended to travel. Also we can know the percentage of Delay and Not delayed of a particular journey by entering Source , Destination and the name of Airlines.

PROJECT OBJECTIVES

1. Given certain features, we would predict that whether a flight would be delayed or not.
2. Also we can predict, how far the delay would be in minutes.
3. We can also predict the percentage of delayed and non-delayed by giving the source, Destination and name of the Airlines.
4. Another feature was added to rank airlines and to recommend which airlines to prefer for a journey.

ARCHITECTURE DIAGRAM



MODULE DESCRIPTION

The dataset was located from Kaggle, this dataset was collected from U.S department of transportation. The dataset tracks the performance of domestic flights within the united states. Handling missing values – The dataset contains small percentage of missing values for certain columns like Departure delay, taxi out and so on. These rows containing missing values are dropped as they make up a very small portion of the dataset .

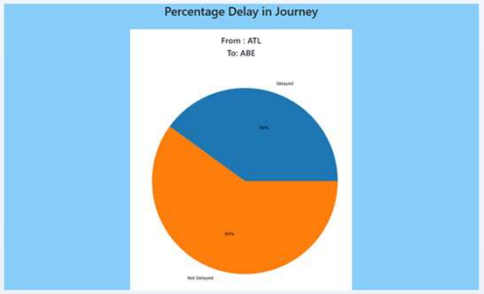
Exploratory Data Analysis (EDA) is done to uncover relationships between different variables. It is followed by feature selection which helps to find best set of features that allows one to build useful models of studied phenomena. So here certain features like Airline, Origin Airport, Destination Airport, Distance, Departure Delay, Scheduled Time, Airtime, Taxi Out are selected for prediction of flight delay.

At First prediction of whether the flight will get delayed or not is done, this is useful for the airline management system to know how far these features can be altered to avoid delays. Also the percentage of delayed to non delayed can be predicted for a particular flight for which ,the origin airport name, destination airport name and airlines name are got as input from the user, based on which other features are extracted from the dataset and finally the user gets a pie-chart of percentage of delayed to non delayed status.

Using regression techniques delay in minutes can be predicted which explains the changes in criterions in relation to changes in select predictors. This can be beneficial to airline management since they have all the features needed to prediction.

Finally ,for a origin and destination , the airways are grouped by their company and for each airline company the score is calculated accordingly. Then the airlines are sorted according to their scores and recommendations are given to the user, based on which the user can select the airlines to book for his journey. These machine learning predictions are then deployed using Django created for user and airline transport system to get delays of flights and recommendations

SCREENSHOTS



Contact:
St. Joseph's College of Engineering
OMR, Semmanchery,
Chennai 600119

Guide:
Mr. S. Vinu M.E., (Ph.D.)