

CENG 621 Machine Learning
Programming Assignment

This assignment is about predicting delays in passenger flights.

Use <https://www.kaggle.com/giovamata/airlinedelaycauses/data#> to download the training data. The data includes more than 1,000,000 records where each record has 30 columns (features).

Choose relevant numeric features to create a regression model to predict how much delay there will be in a flight.

You may combine different delay columns to define the output of your model (or choose the more appropriate one)

Put aside the records with missing data (data cleansing)

Use 80% of data for training your model

Then use remaining 20% for testing the model

Find the average error of your model (use testing data only)

You may use Matlab or R to implement your model

Submit your code and a report before April 23, 2020