Specially Crafted by

Insan Ramadhan

Home Credit
Data Science Bootcamp

Practice Case 04

Updated
Q4.2019

Foundation of Machine Learning

All the task below will be answered using the given dataset which is provided in the e-Learning System
Dataset is a historical flight data in Malaysia from Oct 2018 – Nov 2018
Data consist of information of flight such as:

1. Date of flight
2. Date of arrival
3. Departure Delay
4. Tail Number
5. Airline Name
6. Departure Location
7. Arrival Location
8. Flight Number
9. Delay

You can add any outsource data to help your task, but you need to put the source of the data in the glossary

Foundation of Machine Learning

There are 3 models that you have to make:

1. Linear Regression
2. Logistic Regression
3. 1 Supervised model you choose

Task

1. Provide travel recommendation to passenger (best time to take a flight or best airline), so they will not get caught in delay. Each recommendation has to be supported by at least 1 graph
2. Create a model to estimate the delay duration (Linear Regression)
3. Create a model to predict delay > 60 min (Logistic regression and another supervised model you choosing)
4. Did you do some feature engineering on the dataset ? if yes, please give the reason for each feature you created
5. Using those models predict delay that will happened in December for delay >60

Foundation of Machine Learning

All the results will be put in jupyter notebook format

The jupyter notebook format has to be in a format that cover every step in machine learning model creation step, from importing the dataset until model evaluation

Submit the link to your jupyter notebook in the Practice Case 04 - Submission

Good luck and enjoy your test!