Machine Learning Nanodegree Engineer

Capstone Project

Bharat Singh Jan-19



Flight Delay Prediction

Table Contents

Definition	2
Project Overview	2
Problem Statement	2
Matrices	2
Precision	Error! Bookmark not defined.
Analysis	4
Data Exploration	4
Exploratory Visualization	4
Algorithms and Techniques	4
Benchmark	4
Methodology	4
Data Pre-processing	4
Implementation	4
Refinement	4
Results	4
Model Evaluation and Validation	4
Justification	4
Conclusion	4
Free-Form Visualization	4
Reflection	4
Improvement	4
Quality	4
Presentation	4
Functionality	4
References	4

Definition

Project Overview

This project has been initiated form airline domain. The main objective of this **Flight Delay Prediction** Machine learning project is to predict aircraft delay. It will help in resource management of airline industry. It will tell, particular aircraft will be delay or will reach on time on destination airport. So by using prior knowledge, the can manage resource. Like ground staff, taxi and baggage etc.

This problem comes under the supervised Classification Problem.

Dataset:

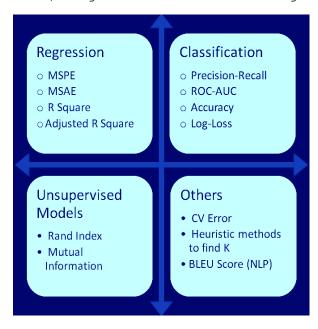
https://www.kaggle.com/fabiendaniel/predicting-flight-delays-tutorial/data

Problem Statement

This is the supervised learning problem, so we will use classification algorithm. It will classify, that given flight will be delay or not.

Matrices

For validation of machine learning model preformation on unseen data or for verifying, system have generalized well for unseen data or not. We need some evaluation matrices. Machine Learning have different matrices for different type of problem. Like Classification, regression or clustering

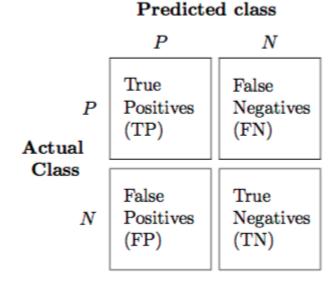


As our problem is related to Supervised Classification Machine learning, we will use all the classification related to matrices.

Confusion Matrix

Confusion matrix is a table representation of model output, which is used to validate the classification model preformation on set of

testing data for which resultant values are known. It also required to calculate Precision, Recall, Accuracy and AUC-ROC Curve.



https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

Accuracy

Accuracy is a metric, which is required to check the model accuracy on the unseen data.

https://developers.google.com/machine-learning/crash-course/classification/accuracy

Recall

Recall is an evaluation matric for Machine Learning Classification Model. What's the probability of correctly classifying for the given positive sample?

Out of all the positive classes, how much model predicating correctly. It should be high as possible.

Precision

ROC-AUC

Analysis

Data Exploration

Exploratory Visualization

Algorithms and Techniques

Benchmark

Methodology

Data Pre-processing

Implementation

Refinement

Results

Model Evaluation and Validation

Justification

Conclusion

Free-Form Visualization

Reflection

Improvement

Quality

Presentation

Functionality

References

https://developers.google.com/machine-learning/crash-course/classification/accuracy

- https://www.kaggle.com/c/flight-delays-spring-2018/data
- https://medium.com/greyatom/performance-metrics-for-classification-problems-in-machine-learning-part-i-b085d432082b
- https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/
- https://scikit
 - learn.org/stable/modules/generated/sklearn.metrics.confusion matrix.html
- <a href="https://www.quora.com/What-is-the-difference-between-the-project-background-problem-definition-aims-of-project-project-justification-and-scope-of-project-project-background-problem-definition-aims-of-project-project-justification-and-scope-of-project-background-problem-definition-aims-of-project-project-justification-and-scope-of-project-background-problem-definition-aims-of-project-project-justification-and-scope-of-project-background-problem-definition-aims-of-project-project-justification-and-scope-of-project-background-problem-definition-aims-of-project-project-justification-and-scope-of-project-background-problem-definition-aims-of-project-project-justification-and-scope-of-project-background-problem-definition-aims-of-project-project-justification-and-scope-of-project-background-problem-definition-aims-of-project-project-justification-aims-of-project-background-project-background-project-
- https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62
- https://en.wikipedia.org/wiki/Precision and recall
- https://www.kaggle.com/c/flight-delays-fall-2018/kernels
- https://www.kaggle.com/c/flight-delays-prediction
- https://machinelearningmastery.com/data-leakage-machine-learning/

• https://scikitlearn.org/stable/auto examples/ensemble/plot forest importances.html