

Literature Review

Team: IBM-Project-10948-1659246792

1. Airline Flight Delay Prediction Using Machine Learning Models

Authors: Yuemin Tang

Date of Publication: October 2021

Flight delays are gradually increasing and bring more financial difficulties and customer dissatisfaction to airline companies. To resolve this situation, supervised machine learning models were implemented to predict flight delays. The data set that records information of flights departing from JFK airport during one year was used for the prediction. Seven algorithms (Logistic Regression, K-Nearest Neighbor, Gaussian Naïve Bayes, Decision Tree, Support Vector Machine, Random Forest, and Gradient Boosted Tree) were trained and tested to complete the binary classification of flight delays. The evaluation of algorithms was fulfilled by comparing the values of four measures: accuracy, precision, recall, and f1-score. These measures were weighted to adjust the imbalance of the selected data set. The comparative analysis showed that the Decision Tree algorithm has the best performance with an accuracy of 0.9777, and the KNN algorithm has the worst performance with an f1-score of 0.8039. Tree-based ensemble classifiers generally have better performance over other base classifiers.

2. Prediction of Flight Time Deviation for Lithuanian Airports Using Supervised Machine Learning Model

Authors: Pavel Stefanovic, Rokas Strimaitis, and Olga Kurasova

Date of publication: 27 December 2020

Machine learning (ML) and deep learning are subareas of artificial intelligence. The main aim of ML is the practice of using algorithms to analyze data, learn from

results, and then make a determination or prediction about something in the world. For example, the machine learning model is trained to predict the weather according to the past weather information. There are many different types of machine learning algorithms [21], and they are always constantly updated or modified, but usually, they are grouped by learning style: supervised learning, unsupervised learning, and semisupervised learning. Also, ML models can be grouped on similarity in form or function such as classification, regression, clustering, prediction, and deep learning.

In the paper, the supervised machine learning model has been used. The general scheme of the used model is presented in Figure 1. First of all, the model has been trained using a dataset according to the past flight information. The dataset is always updated, so the model always is retrained. In the data preprocessing node, the dataset have to be prepared correctly for each algorithm independently, because some of them have an issue with different variables type. Each algorithm is evaluated using cross-validation.

3. Airline delay prediction by machine learning algorithms

Authors: H. Khaksar, A. Sheikholeslami

Date of publication: 28 January 2017

Flight planning, as one of the challenging issue in the industrial world, is faced with many uncertain conditions. One such condition is delay occurrence, which stems from various factors and imposes considerable costs on airlines, operators, and travelers. With these considerations in mind, we implemented flight delay prediction through proposed approaches that are based on machine learning algorithms. Parameters that enable the effective estimation of delay are identified, after which Bayesian modeling, decision tree, cluster classification, random forest, and hybrid method are applied to estimate the occurrences and magnitude of delay in a network. These methods were tested on a U.S. flight dataset and then refined for a large Iranian airline network. Results showed that the parameters affecting delay in US networks are visibility, wind, and departure time, whereas those affecting delay in Iranian airline flights are fleet age and aircraft type. The proposed approaches exhibited an accuracy of more than 70% in calculating delay occurrence and magnitude in both the whole-network US and Iranian. It is hoped that the techniques put forward in this work will enable airline companies to accurately predict delays, improve flight planning, and prevent delay propagation.

4. Using Big Data-machine learning models for flight delays analytics

Authors: Thérence Nibareke & Jalal Laassiri

Date of publication: 17 September 2020

Nowadays large data volumes are daily generated at a high rate. Data from health system, social network, financial, government, marketing, bank transactions as well as the sensors and smart devices are increasing. The tools and models have to be optimized. In this paper we applied and compared Machine Learning algorithms (Linear Regression, Naïve bayes, Decision Tree) to predict diabetes. Further more, we performed analytics on flight delays. The main contribution of this paper is to give an overview of Big Data tools and machine learning models. We highlight some metrics that allow us to choose a more accurate model. We predict diabetes disease using three machine learning models and then compared their performance. Further more we analyzed flight delay and produced a dashboard which can help managers of flight companies to have a 360° view of their flights and take strategic decisions.