LITERATURE SURVEY

REFERENCE PAPERS :

Reference paper : 1

Topic : Airline Flight Delay Prediction Using Machine Learning Models, 2021 5th International Conference on E-Business and Internet, Singapore, Singapore, October 2021.

Description :

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 Neighbour, Gaussian Naïve Bayes, Decision Tree, Support Vector Machine, 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. 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. Tree-based ensemble classifiers generally have better performance over other base classifiers.

Reference Link : https://dl.acm.org/doi/fullHtml/10.1145/3497701.3497725

Reference paper : 2

Topic : Flight delay prediction based on deep learning and Levenberg-Marquart algorithm, 26 November 2020.

Description :

Flight delay is inevitable and it plays an important role in both profits and loss of the airlines. An accurate estimation of flight delay is critical for airlines because the results can be applied to increase customer satisfaction and incomes of airline agencies. However, most of the proposed methods are not accurate enough because of massive volume data, dependencies and extreme number of parameters. This paper proposes a model for predicting flight delay based on Deep Learning (DL). DL is one of the newest methods employed in solving problems with high level of complexity and massive amount of data. Moreover, DL is capable to automatically extract the important

features from data. Furthermore, due to the fact that most of flight delay data are noisy, a technique based on stack denoising auto-encoder is designed and added to the proposed model.

Reference Link :

https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00380-z

Reference paper : 3

Topic : Machine Learning Model - based Prediction of Flight Delay

Description :

Prior prediction of flight arrival delays is necessary for both travelers and airlines because delays in flights not only trigger huge economic loss but also airlines end up losing their reputation that was built for several years and passengers lose their valuable time. Our paper aims at predicting the arrival delay of a scheduledindividual flight at the destination airport by utilizing available data. The predictive model presented in this work is to foresee airline arrival delays by employing supervised machine learning algorithms. US domestic flight data along with the weather data from July 2019 to December 2019 were acquired and are used while training the predictive model. XGBoost and linear regression algorithms were applied to develop the predictive model that aims at predicting flight delays.. Flight data along with the weather data was given to the model. Using this data, binary classification was carried out by the XGBoost trained model to predict whether there would be any arrival delay or not, and then linear regression model predicts the delay time of the flight.

Reference Link : https://ieeexplore.ieee.org/document/9243339

PROPOSED METHODOLOGY :

Using a machine learning model, we can predict flight arrival delays. The input to our algorithm is rows of feature vector like departure date, departure delay, distance between the two airports, scheduled arrival time etc. We then use decision tree classifier to predict if the flight arrival will be delayed or not. A flight is considered to be delayed when difference between scheduled and actual arrival times is greater than 15 minutes. Furthermore, we compare decision tree classifier with logistic regression and a simple neural network for various figures of merit.