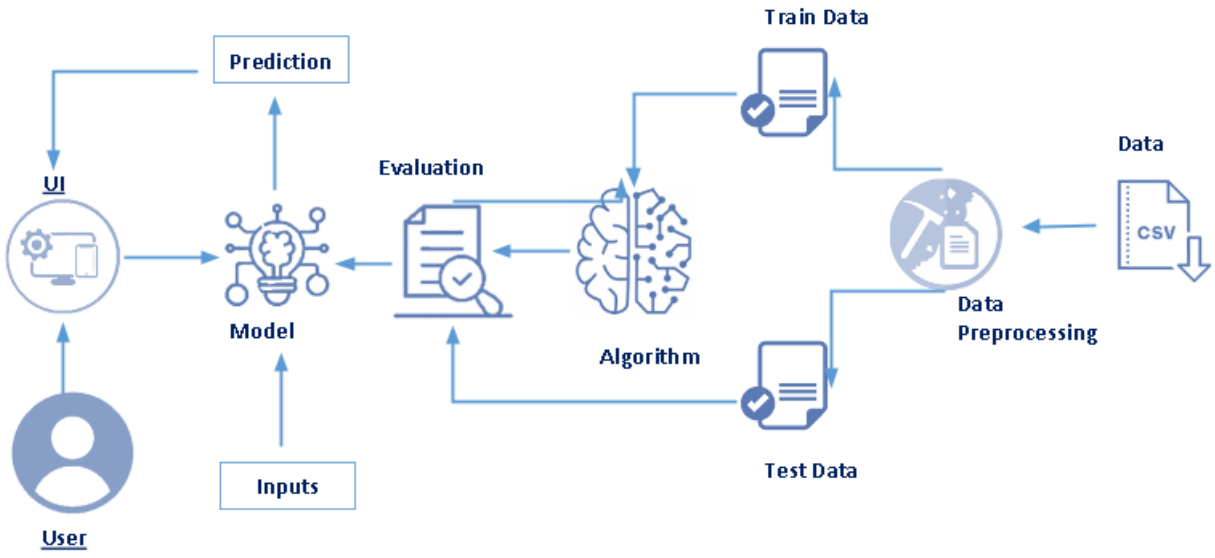
Developing A Flight Delay Prediction Model Using Machine Learning :-

Literature Survey

1 Introduction

Air travel has become widely common and preferred among travelers over the years due to its comfort and the time of travel.This has in a way led to a lot of air traffic and on ground and hence resulting in massive levels of aircraft delays in the air and on ground.The delays are a cause of environmental and economic losses.The proposed model helps to predict flight delay to optimize flight operations and minimize delays in a most accurate manner.machine Learning models can be used to do this.A decision tree classifier can be used to predict if the flight arrival will be delayed or not depending on an input vector. Furthermore, we can compare the decision tree classifier with logistic regression and a simple neural network for various figures of merit.



2 Literature Study

2.1 Research Paper

Yuemin Tang from University of Southern California proposed a paper with the main goal to compare the performance of machine learning classification algorithms when predicting flight

delays. The data set used for analysis contains data about flights leaving from JFK airport between one year from November 2019 to December 2020. In this study, classification models were selected and trained using seven algorithms: Logistic Regression, K-Nearest Neighbor (KNN), Gaussian Naïve Bayes, Decision Tree, Support Vector Machine (SVM), Random Forest, and Gradient Boosted Tree. The value of each evaluation measure for every algorithm is presented in table as result.The result shows that decision Tree performs well when predicting flight delays in the data set. Other tree-based ensemble classifiers also show good performance. Random Forest and Gradient Boosted Tree have an accuracy of 0.9240 and 0.9334, significantly higher than the rest of the models. The other four base classifiers Logistic Regression, KNN, Gaussian Naïve Bayes, and SVM, are not tree-based and did not show good performance. The KNN model is the worst performed since its precision and f1-score are the lowest among the seven models.

Reference:

Yuemin Tang. 2021. Airline Flight Delay Prediction Using Machine Learning Models. In *2021 5th International Conference on E-Business and Internet (ICEBI 2021), October 15-17, 2021, Singapore, Singapore*. ACM, New York, NY, USA, 7 Pages.
https://doi.org/10.1145/3497701.3497725

2.2 Related Research Work

Bhuvan Bhatia used Python based Logistic Regression
along with Support Vector Machine to predict flight delays and compared the results with other models such as Random
Forest Classifier and derive the best classifier to solve the problem. The dataset focused on LaGuardia International Airport. The result shows that the Random Forest method yields the best performance compared to the SVM model.

Link :

https://scholarworks.calstate.edu/