

# **PREDICTING FLIGHT DELAYS WITH ERROR CALCULATION USING MACHINE LEARNED REGRESSION**

## **ABSTRACT**

Flight delay is a major problem in the aviation sector. During the last two decades, the growth of the aviation sector has caused air traffic congestion, which has caused flight delays. Flight delays result not only in the loss of fortune but also negatively impact the environment. Flight delays also cause significant losses for airlines operating commercial flights. Therefore, they do everything possible in the prevention or avoidance of delays and cancellations of flights by taking some measures. In this paper, using machine learning models such as Logistic Regression, Decision Tree Regression, Bayesian Ridge, Random Forest Regression and Gradient Boosting Regression we predict whether the arrival of a particular flight will be delayed or not.

## **EXISTING SYSTEM ALGORITHMS**

Random forest, Gradient Boosting Classifier, Support Vector Machine and the k-nearest neighbour algorithm, where 77% / 79.7% of accuracy is maximum for the above used algorithms.

Datasets are available in UCI respo. Or Kaggle.com

## **PROPOSED ALGORITHMS**

We are going to use Decision Tree Regression, Ridge Regression, Logistic Regression, Random Forest Regression, Gradient Boosting Regression with the existing algorithms for accuracy comparison

## **LITERATURE SURVEY**

As discussed, considering the standard taxonomy of the flight delay and its problems, one will contemplate the scope of prediction to be one in every of these factors or combination of those factors. The models developed during this system may be applied to predict the incidence of flight delay at airports. Such prognostication capabilities would facilitate traffic managers and airline dispatchers to organize mitigation methods for reducing traffic disruptions. This issue can be reduced by developing the flight delay prediction tool which can be developed using the following methods.

Statistical analysis Statistical model requires the use of correlation analysis, parametric and non parametric tests, multivariate analysis and econometric models. Government agencies have invested in these econometric models to understand the relationship between delay and Passenger demand, fare, size of aircraft etc . Probabilistic models Probabilistic models require analysis tools that estimate the probability of an event based on the historic data. The estimated outcome is given in the form of a distribution function of the probability. The factor of randomness

- Classification – It is a type problem in which the output variable is an entire category itself, such as “Win” or “Lose”, the entire input data is classified into the category variables; it is generally used largely for recommendation problems
- Regression – It is a type of problem in which the output variable is a real value, such as few raw data values related to something.