**Flight Delay Prediction Based on Aviation Big Data and Machine Learning**

**ABSTRACT:**

Accurate flight delay prediction is fundamental to establish the more efficient airline business. Recent studies have been focused on applying machine learning methods to predict the flight delay. Most of the previous prediction methods are conducted in a single route or airport. This paper explores a broader scope of factors which may potentially influence the flight delay, and compares several machine learning-based models in designed generalized flight delay prediction tasks. To build a dataset for the proposed scheme, automatic dependent surveillance broadcast (ADS-B) messages are received, pre-processed, and integrated with other information such as weather condition, flight schedule, and airport information. The designed prediction tasks contain different classification tasks and a regression task. Experimental results show that long short-term memory (LSTM) is capable of handling the obtained aviation sequence data, but overfitting problem occurs in our limited dataset. Compared with the previous schemes, the proposed random forest-based model can obtain higher prediction accuracy (90.2% for the binary classification) and can overcome the overfitting problem.

**EXISTING SYSTEM:**

* Y. J. Kim et al. proposed a model with two stage. The first stage is to predict day-to-day delay status of specific airport by using deep RNN model, where the status was defined as an average delay of all flights arrived at each airport.
* The second stage is a layered neuron network model to predict the delay of each individual flight using the day-to-day delay status from the first stage and other information. The two stages of the model achieved accuracies of 85% and 87.42%, respectively.
* This study suggested that the deep learning model requires a great volumes of data. Otherwise, the model is likely to end up with poor performance or overfitting

**DISADVANTAGES OF EXISTING SYSTEM:**

* Several reasons are restricting the existing methods from improving the accuracy of the flight delay prediction. The reasons are summarized as follows: the diversity of causes affecting the flight delay, the complexity of the causes, the relevancy between causes, and the insufficiency of available flight data.
* The air route information (e.g., traffic flow and size of each route) was not considered in their model, which prevents them from obtaining higher accuracy.

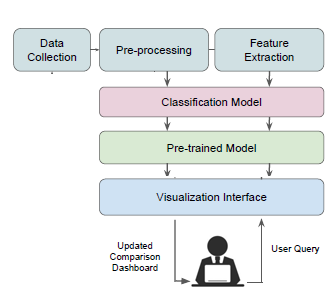
**PROPOSED SYSTEM:**

* We explore a broader scope of factors which may potentially influence the flight delay and quantize those selected factors. Thus we obtain an integrated aviation dataset. Our experimental results indicate that the multiple factors can be effectively used to predict whether a flight will delay.
* Several machine learning based-network architectures are proposed and are matched with the established aviation dataset. Traditional flight prediction problem is a binary classification task. To comprehensively evaluate the performance of the architectures, several prediction tasks covering classification and regression are designed.
* Conventional schemes mostly focused on a single route or a single airport. However, our work covers all routes and airports which are within our ADSB platform.

**ADVANTAGES OF PROPOSED SYSTEM:**

* Our work benefits from considering as many factors as possible that may potentially influence the flight delay. For instance, airports information, weather of airports, traffic flow of airports, traffic flow of routes.
* The random forest-based architecture obtained a testing accuracy of 90.2% for the binary classification, which is considered a promising result and demonstrate the strong ability of the ensemble learning

**SYSTEM ARCHITECTURE:**



**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Pentium i3 Processor.
* Hard Disk : 500 GB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 2 GB

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows 10.
* Coding Language : Python

**REFERENCE:**

Guan Gui, Senior Member, Fan Liu, Student Member, Jinlong Sun, Jie Yang, Ziqi Zhou, and Dongxu Zhao, “Flight Delay Prediction Based on Aviation Big Data and Machine Learning”, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, 2020.