**LITERATURE SURVEY**

INTRODUCTION

Flight delay prediction is fundamental to establish the more efficient airline business. The development of accurate prediction models for flight delays became cumbersome due to the complexity of air transportation system, the number of methods for prediction, and the deluge of flight data. In this context, this paper presents a thorough literature review of approaches used to build flight delay prediction model. Airlines delays make immense loss for business field as well as in budget loss for a country. Flight delays hurt airlines, airports, and passengers. We are proposing machine learning algorithms like Linear regression Techniques. The aim of this research work is to predict Flight Delay, Which is highest economy producing field for many countries and among many transportation this one is fastest and comfort, so to identify and reduce flight delays, can dramatically reduce the flight delays to saves huge amount of turnovers, using machine-learning algorithms. Flight delays could always be annoying, especially in the case when the period of delay was so long that there was even a danger to miss the next flight. However, if there was a way to predict whether there would be a delay or even better – how long the delay could be, then people could make earlier preparation to reschedule following flights in an earlier manner.

LITERATURE REVIEW

**In the year of 2021, Rajkumar Goel Institute of Technology**

After applying both the models for predicting whether a flight should be delayed, as well as how much one would expect a flight should be delayed, they found the following factors to be important: week, month, airline carrier reference, planned elapsed time (in airtime), distance between two departure and destinations, flight planned departure time, departure airport code, and taxi-in and taxi-out4 time. By applying there model, on the data collected, one could be able to predict whether a flight might be delayed, and more importantly, how long delayed time she/he would expect. However, there is some limitation in our model, first, our model only included one-year data due to our computation capability, as more years of data included, the prediction could be easier. In addition, some other related information such as airplane type, e.g., detailed weather data specific to airport were not included. Therefore, researchers could try to collect more related data and deploy better computational powers to build a better model. This paper presented a methodology for predicting aggregate flight departure delays in airports by exploring supervised learning methods. This way, they may be able to predict the delays of a new flight, without needing several months of data to build a prediction model. Another step forward would be to generalize the model to flights of the entire world, or at least to exploit more data sources, to build more complete predictions. Finally, the most interesting step would be to integrate such a model into a flight booking tool, to provide the delay prediction to future passengers, even this would require a strong confidence in the information provided, considering the possible impact in terms of reservations

**In the year of 2020, Nanjing University of Aeronautics and Astronautics**

This paper presented a methodology for predicting aggregate flight departure delays in airports by exploring supervised learning methods. The proposed new model was enabled by four types of airport-related aggregate characteristics, including time characteristics, flight plan characteristics, delay characteristics and local weather characteristics. The results obtained show that for a 1-h forecast horizon, LightGBM model provides the best result, giving 0.8653 accuracy with 6.58 min mean absolute error, which is 1.83 min less than previous research. Analysis also found that accumulated number of departure demand in the prediction period is the dominating factor in the LightGBM model. The number of planned departures in the prediction period and the expected delay time of departures before the prediction period are two other obvious factors, while the expected delay time of arrivals before the prediction period, hour of the day, and the number of planned departures before the prediction period are three following characteristics. Of special note is that the model performances with local weather characteristics are not as good as those without meteorological data. Two potential reason are that the cancelled and returned flight records caused by local weather characteristics can hardly be translated into specific delay time in departing airports, and the local flight delays are often caused by weather conditions en-route or in the airports of the previous flights, not just in local airports. The prediction model presented in this paper yields a better understanding of delays interactions between time, flight plan and previous delay. Since we predict the flight departure delays from the airport aspect, the model could be used for reminding airport managers, air traffic controllers and passengers to deal with the impending congestion in airports. Future works include exploring some other explanatory characteristics such as national weather, city-pair, and network states, etc., and extending forecast horizon with more accuracy results.

**In the year of 2019, H. Khaksar and A. Sheikholeslami,Sharif University of Technology**

FDP methods, namely decision tree, cluster, Bayesian, random forest, and hybrid classication, were proposed in this research. These approaches were examined on the basis of real datasets on US and Iranian ight networks. The results indicated that the hybrid approach exhibited a performance superior to those of the other methods and was therefore adopted as the FDP model. Parameters such as eet age and aircraft type exert strong eects on ight delays in the Iranian network, whereas weather conditions strongly in uence ight delays in the US network. The accuracy levels of the hybrid approach were 71.39% and 76.44% in predicting delay occurrence and 70.16% and 75.93% in predicting delay magnitude in the US and Iranian networks, respectively. These results may be of interest to airlines that want to implement measures for preventing delay propagation, especially those based in developing countries, such as Iran. For the future studies, researchers can implement other exciting data mining methods and compare the results. The proposed combined method of delay anticipation and its results can also be further explored in other studies. For example, combing the hybrid method with robust ight scheduling shows potential as an interesting research direction.

**In the year of 2017 , Anish M. Kalliguddi\* , Aera K. Leboulluec,College of Engineering, University of Texas, United States**

This study is devoted to develop a predictive model to forecast flight delays. Data spanning for over 1million observations including US domestic flights variables was used. Models based on multiple linear regression, decision trees and random forest algorithms are created and tested in R-studio software concluding that Random forest model outperforms other two models based on the evaluation criteria. In addition, the study also sheds light on the significant factors responsible for departure delay. The splitting variables or the significant variable are found to be late aircraft delay, Carrier delay, weather delay and NAS delay which have the most effect on on-time flight departure. The predictive model was developed for a period of 365 days for all US domestic airports. It is seen that the longer forecast horizon helps in better prediction accuracy with minimum prediction error for random forest. These models can be used to improved traffic management decision in comparison with the current applications of Enhanced Traffic Management Systems (ETMS). Although the model gives very good prediction accuracy, more variables can be considered to develop a predictive model. For example, Weather data can be extracted and used to better develop a predictive model for flight delay. The future scope of this study involves various approaches that can be used to analyze the data. Principal component analysis or transformation can be done to uncover hidden relations between variables. In addition, since the data is not exactly linear, artificial neural networks or Support vector machines can be used to analyze the effect of various variables on flight delay.

**In the year of 2016, Conference Organized by Missouri University of Science and Technology**

A new ANN structure (DMP-ANN) is introduced which is suitable for prediction of defects such as delays in operations. This structure is appropriate for problems with nominal variables, where traditional ANN models have difficulties. For example, the types of cargo or ID number of origin of departure are variables that cannot be directly used in a traditional ANN. The input layer in proposed DMP-ANN consists of several sublayers in which one or more neurons are active (output=1) and others where they are inactive (output=0). Hence, the learning process involves updating the weights of active neurons. The introduced ANN model is applied to a system of airport traffic control where the arriving flights are prioritized for landing based on the expected possible delays. The results suggest that the proposed method can be effective for specific problems that include many nominal variables, such as the transportation problem. One of the limitations of this study that needs to be addressed in our future work is the complexity of the proposed method (as the number of variables increases the number of connections also significantly increase). Furthermore, we will consider the integration of the proposed method with fuzzy logic to expand the real-world applications of the proposed method.

CONCLUSION:

The studies mentioned above offer some suggestions for contemporary approaches to tracking expenses. These kinds of studies frequently demonstrate how ideas change over time. Evolution is not a prerequisite; rather, it's a shift in thinking and time, during which we estimate, evaluate, and assess things in accordance with changing needs. Although there are some android apps, the technology used in these projects was similar to that used in earlier times. Nevertheless, these apps have their own set of problems. Additionally, I believe that things need to be much simpler to use on a desktop device. Because Android apps sometimes produce accurate results.