Team 094 Final Report

Hira Tanweer, Hymee Huang, Kashif Naseem, Neha Rani Singh, Umair Zakir Abowath

Introduction:

Currently, the commercial airline industry relies on simple statistics about on-time arrival/departures that are sometimes provided to passengers alongside their airline reservation. In addition, certain internet based real time flight tracking services provide the same statistical data [1]. Some flight delay information can be derived indirectly from FAA's National Airspace System which provides information on airport events in real time [2]. At present, there appear to be no commercially available services which are able to provide prediction or probability of delay for a specific flight. There is no commercial usage of any ML models in the airline industry to assist consumers in predicting flight delays.

Problem definition:

Our objective is to provide highly transparent delay prediction information to the air travelers in the US. To accomplish this, we propose to develop a visualization application that leverages the power of machine learning to predict domestic flight delays in the country. The application will feature an interactive user interface enabling the user to furnish flight particulars and obtain an estimate of the expected delay. This objective is critical since customers anticipate three levels of service quality: desired, adequate, and predicted [3], where customers not only want to know if their schedule is delayed, but also want to know the anticipated estimation on their schedule.

Literature Survey:

In academia, there has been considerable work done on building and applying ML based models to analyze flight delays. Some studies have focused on certain specific aspects of aviation to predict delays. For example, use of Markov chain algorithm to model some processes (like baggage handling) which are critical to flight punctuality and then using these models to predict flight delays [4] or focusing on meteorological data and using Dual-channel Neural Networks to predict delays caused by weather factors [5]. These were useful studies but were lacking in that they did not wholistically look at flight and weather specific information together in their models. Other studies utilized a much wider array of parameters to model aircraft delay and used a variety of modeling techniques including but not limited to deep learning and neural networks ([6], Random Forest [7], Support Vector ([8], LinearR, ExtraRT ([9], extreme gradient boosting algorithms ([10] and Stacking algorithms [11]. Some studies expanded their datasets available for modeling by incorporating communication and surveillance data from air traffic management systems [12]. Although these studies were more detailed, as discussed later, they are not consumable commercially at their current stage. Some experts used as many as 10 ML algorithms to model and evaluate flight operations data and identify the features most suited for optimum prediction accuracy for each model [13]. One study by Stanford on pre-flight data compared Decision Tree, Logistic Regression and Neural Network and was able to show that all 3 of the techniques were able to produce comparable accuracies. The limitations observed was on the training data size [14]. This is a useful result for us because it emphasizes the importance of the datasets rather than the actual modeling method used.

Considerable non-commercial work has been done on the back end to build ML models for flight delay prediction. However, none of these are accessible to general consumer because they are lacking in one crucial aspect which is a user-friendly front end where a consumer or future traveler can navigate through an interactive interface and access flight delay prediction or forecast. Thus, even if adequate ML based models exist to accurately predict flight delays, they are inaccessible to an average consumer.

Proposed Method:

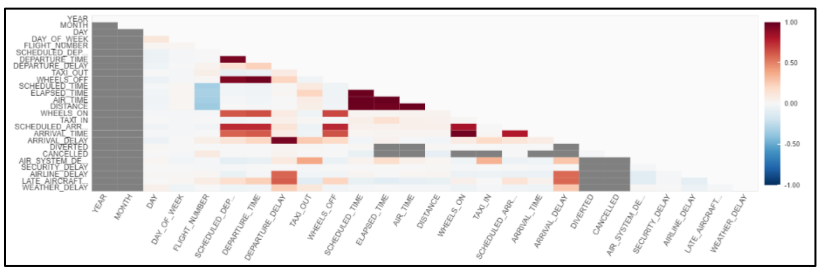
Our high-level solution approach is to use flight and weather datasets to train a ML model to predict flight delay. This model is made accessible to a consumer through an iOS based interactive front end. We used python for tasks pertaining to feature selection, feature analysis and building models. Flight and weather-related datasets were used for data exploration and to identify essential features. The data was then split into training and test sets to fit and evaluate multiple models such as decision trees and linear regression. Evaluation metrics such as accuracy, recall, f1 score, and precision were used to evaluate the performance of the models. Our final modeling approach was to divide flight delays into multiple categories and apply Random Forest Classifier to a multicategory classification, enabling us to predict delay duration. The most innovative aspect of our approach is the creation of an iOS based front end for the model to present delay prediction information to travelers in a transparent and comprehensible manner. The GUI enables travelers to input their flight information and receive output information regarding their flight. This provides a level of accessibility to the consumer not available before [15].

Experiments/Evaluation:

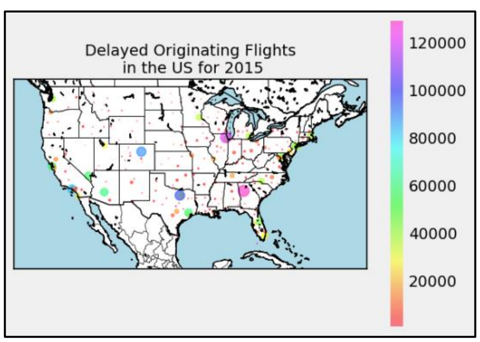
Exploration Data Analysis:

We started our phase of experiments and evaluations with some basic data exploration. We first analyzed our data at a high level looking at some basic statistical parameters for example, correlation matrix between variables.

Dataset Statistics	
Number of Variables	31
Number of Rows	165843
Missing Cells	783826
Missing Cells (%)	15.2%
Duplicate Rows	0
Duplicate Rows (%)	0.0%
Total Size in Memory	76.3 MB
Average Row Size in Memory	482.6 B
Variable Types	Categorical: 10 Numerical: 21



The next step in our data exploration was to look at some individual variables that we think are important and their impact on total flight delay separately from other variables.



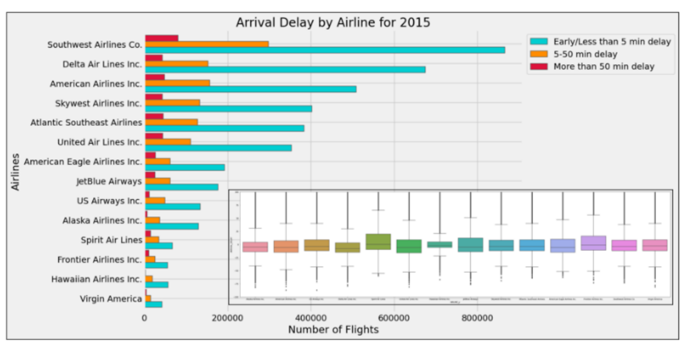


Figure 2a) Arrival delays by airport of origin *2b) Arrival delays by airline*

For example, if we look at the average daily delay throughout the year of 2015 with date, we can see some trends (Figure 3a). We decided to analyze at the 7 day (weekly) rolling averge of the daily delays to reduce noise. Clear spikes can be observed around Dec/Jan, Feb/March and Nov/Dec. These correspond to the Christmas holidays, spring break and thanksgiving. In addition, we can also observe a series of spikes between Jun-Aug which corresponds to summer vacations for schools in the US.

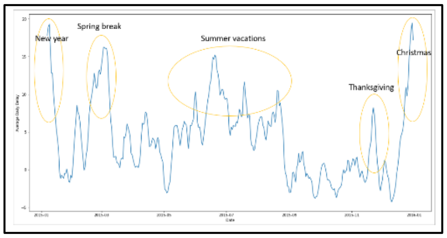
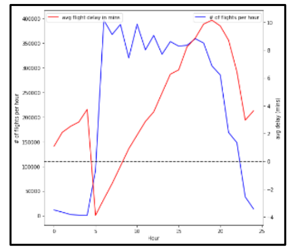
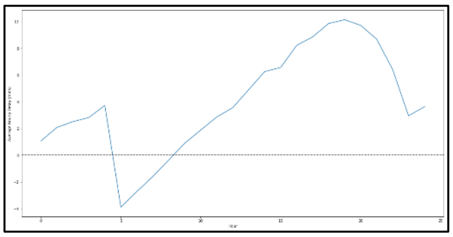


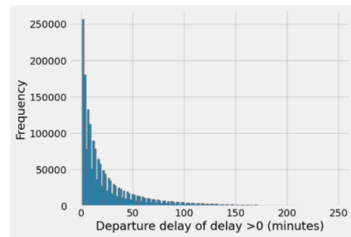
Figure 3a) 7 day (weekly) rolling average of the daily delays vs date



3b)Daily average delays vs hour of departure



3c) Daily average delays vs # of flights in a hour



3d) Frequency of delays per bucket of delay time

In order to analyze if the time of the day at which a flight is scheduled to depart has any impact on delay, we created a scatter plot of the scheduled departure time vs delay in arrival in mins (Figure 3b). To provide more clarity, we decided to group flights by hours of scheduled departure.
Some patterns can easily be observed. For example, the flights departing between 10 AM and 10 PM are prone to most delays. Flights departing between midnight and 9 AM are mostly on time as the average flight delays are 0 or below. This intuitively makes sense as most US airports are in urban areas and shut down or reduce flight operations at night. This means the air traffic congestion is greatly reduced and with it, flight delays. To further prove this, we plot the number of flights in an hour and the average flight delay on the same plot (Figure 3c). It is interesting to note that the flight delays do not increase or decrease immediately with the increase or decrease with the number of departing flights. There is a lag effect where the flight delays increase gradually and build up as the as the day goes on and more and more flights are added to the air traffic system, increasing the congestion and load on the system.

ML Models:

Initially, we tried to model flight delay as a continuous variable, however, we soon found that our models were inadequate in modeling the variability of the data resulting in extremely low R-squared values. Therefore, we decided to divide our response variable into bins or buckets. These bins will vary in size ranging from a few mins for shorter delays to a few hours in case of longer delays. We decided to use random forest classifier to classify the flights (data points) into one of the many delay categories based on various predictors.
To start off with a simple model, we used a handful of predictors including airline, origin and destination airports, scheduled duration of flight, day of week, month, time of day of the flight and the distance between the two airports to categorize the expected delay. Our RFC model was able to achieve >70% accuracy, precision, recall and f1 score for early or on-time flights. However, these performance metrics for the delayed categories were relatively low.
We evaluated a challenge that was causing us to not predict the delayed cases and the lower value of recall and f1 score. We encountered a significant class imbalance issue in our data due to the high percentage of flights that were on time. To address this, we oversampled the minority class. Our initial models focused solely on flight characteristics such as origin airport, destination airport, time of day, airline, and scheduled duration of flight etc. We later incorporated weather data like cloud cover, humidity, wind speed, and pressure etc to improve our models' accuracy and complexity. Since, using a

regression-based model with “delay time” as a continuous outcome variable did not meet our performance metrics, we analyzed in the case of predicting delays we can predict well on the delayed time category (ON-time, <20mins delay, 20-40mins delay, 40-60mins delay, >60mins delay) and on binary outcome prediction of delay or non- delay. We eventually settled on using a Random Forest Classifier, which achieved our desired results with greater than 70% accuracy. This led us to divide flight delays into multiple categories and apply RFC to a multicategory classification, enabling us to predict delay duration with 60-70% accuracy and had a recall, f1 score of > 60 for the categories of delay. Given the large size of our dataset, we divided the flights into regions and built separate models for inter and intra region flights. We used ANOVA to select the best features for each model, identified the optimal feature set for each flight set and hyper tune our models to find the best model possible.

<u>Building an app as the front-end interface:</u>

While building the actual ML models, we were also looking into finding an appropriate method to provide a front end for these models. We looked at several options including d3 and python-based toolkits (PyQT, wxPython etc) but finally decided to go with an app style front interface & developing an iOS app using Swfit & Xcode that visualizes flight delays through a user-friendly interface. To achieve this, we converted our trained Python model into a CoreML model format, which we integrated into the app. The app uses MapKit to display a map of USA airports with pins and also a list view to select departure and arrival airports. It has a date picker to select the flight departure date and time. Users can input flight information such as flight number, date, source, and destination to obtain the predicted delay for their desired flight. The app will be available for both iPhone and iPad users.

Attached are few screenshots of our mobile app work:

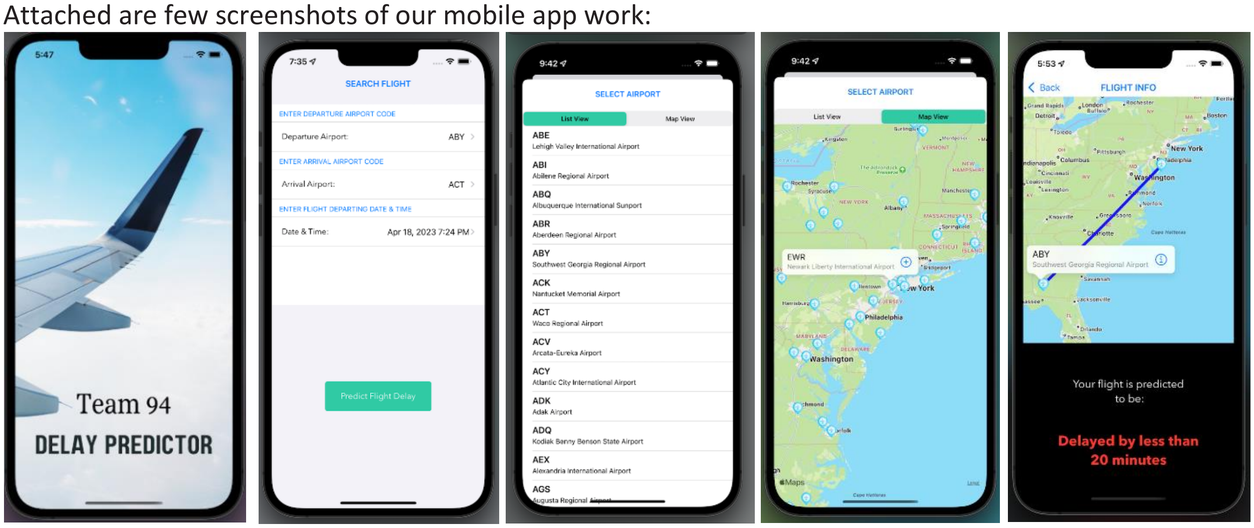


Figure 4a) Delay Predictor Launch screen 4b)List view of airport selection 4c) Map view of airport selection 4d)Interactive map 4e) Output screen

<u>Conclusions and discussions:</u>

From data exploration phase, some predictors like data and time of the flight, airline, origin and destination airports and the distance between them, have a very clear impact on fight delay. We also conclude that delays increase as the number of planes in the air traffic system increases.

In terms of the models, we are able to conclude that RFC based models seem to be much better than others and this is in line with observations in literature. [16]

In evaluating our approaches, we used a combination of statistical metrics such as accuracy, precision, recall and F1-score to assess model performance. We also used cross-validation to ensure the robustness of our models and avoid overfitting. The results of our modeling efforts showed that the Random Forest Classifier was the most effective approach, achieving >70% accuracy in predicting flight delays. We were also able to predict the duration of delays with 60-70% accuracy. Comparing our models to others, we found that our approach outperformed linear and logistic regression in predicting flight delays. Although our modeling approach successfully addressed the class imbalance issue in the data and enabled us to accurately predict delays, we acknowledge that there may be other models or techniques that we did not explore that may yield better results.

One area of improvement and future work on the modeling side can be to try to model delay as a continues variable rather than in bins of 30 mins. Our attempts at doing this did not yield good accuracy so we moved away from this approach, but this will be the something good to pursue agaiin future derivation of this work.

For front end visualization, decided to go with a core ML based iOS app front interface. To aid in the ease of use and level of accessibility. Given the widespread use of smart phones among masses, this approach gives us the best level of accessibility to an average traveler compared to the other options. The result is an interactive, user friendly, and easy to use app that can be accessed and used by anyone.

Logical next step for this app would be to add push notifications. By integrating push notifications with airport and airline data feeds, the app could alert users to potential flight delays or cancellations, giving them valuable information in real time. Additionally, the app could provide personalized recommendations to help users avoid delays or make alternate travel arrangements.

From a development perspective, we could continue to refine and improve the machine learning model powering the app. This might include incorporating additional data sources or optimizing the model to improve prediction accuracy. We could also explore ways to enhance the app's user interface and user experience, adding social media integration so users can share their travel plans, flight delays and experiences on social media platforms. This can also help promote the app and increases user engagement.

Finally, all team members have contributed a similar amount of effort and this project will be prepared for commercial development for members who wish to proceed. Importantly, as pointed out in the HAL Open Science’s paper, "Evaluation should be representative of how the model would perform when deployed in a real-life setting." [17] Therefore, the team tried to incorporate this perspective into the final product to ensure its suitability for real-world applications.

Bibliography

[1] C. Elliott, *These Sites Will Help You Handle Any Airline Cancellation Or Delay,* 2022.

[2] "National Airspace System Status," [Online]. Available: https://nasstatus.faa.gov/.

[3] E. T. N. P. L. L. A. P. D. R. Marina Efthymiou, "The Impact of Delays on Customers’ Satisfaction: an Empirical Analysis of the British Airways On-Time Performance at Heathrow Airport," *Journal of Aerospace Technology and Management,* 2019.

[4] C.-L. Wu, "Inherent delays and operational reliability of airline schedules," *Journal of Air Transport Management ,* 2005.

[5] T. Z. M. Y. J. L. Jingyi Qu, "Flight Delay Prediction Using Deep Convolutional Neural Network Based on Fusion of Meteorological Data," *Neural Processing Letters,* 2020.

[6] V. Venkatesh, A. Arya, P. Agarwal, S. Lakshmi and S. Balana, "Iterative machine and deep learning approach for aviation delay prediction," in *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)*, Mathura,, 2017.

[7] J. Z. a. C. L. Yutong Wang, "The analysis of the influence of delay absorbing sequence on flight delay propogation," *IOP Conference Series: Materials Science and Engineering,* 2019.

[8] S. M. Ehsan Esmaeilzadeh, "Machine Learning Approach for Flight Departure Delay Prediction and Analysis," *Transportation Research Record: Journal of the Transportation Research Board,* p. 145–159, 2020.

[9] B. L. T. L. W. Bojia Ye, "A Methodology for Predicting Aggregate Flight Departure Delays in Airports Based on Supervised Learning," *Sustainability ,* 2020.

[10] Ö. T. a. N. T. Irmak Hatıpoğlu, "FLIGHT DELAY PREDICTION BASED WITH MACHINE LEARNING," *Scientific Journal of Logistics,* pp. 97 - 107, 2022.

[11] H. Z. H. L. Z. a. G. L. Jia Yi, "Flight Delay Classification Prediction Based on Stacking Algorithm," *Journal of advanced transportation,* 2021.

[12] J. S. ,. Y. Z. Z. D. Z. Guan Gui Fan Liu, "Flight Delay Prediction Based on Aviation Big Data and Machine Learning," *IEEE Transactions on Vehicular Technology,* pp. 140-150, 2020.

[13] M. B. M. Ş. V. T. D. K. Mehmet Cemal ATLIOĞLU, "Supervised Learning Approaches to Flight Delay Prediction," *Sakarya University Journal of Science,* 2020.

[14] N. K. a. N. Jamadagni, "Application of Machine Learning Algorithms to," 2017.

[15] K. G. a. H. Balakrishnan, "A Comparative Analysis of Models for Predicting Delays in Air Traffic Networks," in *Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017)*, 2017.

[16] Y. Tang, "Airline Flight Delay Prediction Using Machine Learning Models," in *ICEBI* , Singapore, 2021.

[17] O. C. Gaël Varoquaux, *Evaluating machine learning models and their diagnostic value,* HAL open science.