Industrial Engineering and Applications
L.-C. Tang (Ed.)

doi:10.3233/ATDE230058

Prediction of Aircraft Arrival Delay Due to Weather

Patara CHARNVANICHBORIKARN [a,1] and Nantachai KANTANATHA [b]
[a] *Department of Industrial Engineering*
Faculty of Engineering, Chulalongkorn University
[b] *Department of Industrial Engineering*
Faculty of Engineering, Chulalongkorn University

Abstract. The world has seen a dramatic change in weather patterns caused by global warming. Extreme weather and unpredictable events are occurring more often. As weather is the most unpredictable component of a flight, if there is a slight chance of predictability in what the weather is going to be, it could help with the logistical issues and prevent further delays. Therefore, aircraft arrival prediction could potentially be a solution to improve the efficiency of the operation and the effectiveness of the contingency plans to respond to such events. The objective of this paper was to gain more insights and study the effect of weather on aircraft arrival delay by using random forest classification method and adaboost classification method. The results show that random forest classifier is the most accurate model for the weather delay prediction with the accuracy of 92.98 percent. When comparing F-1 score with adaboost, random forest model also gives higher values for all conditions.

Keywords. Weather, Aircraft, Arrival delay

1. Introduction

1.1. US Airline Industry

Considered one of the largest industries in the world, the US airline industry carries 2.5 million passengers per day to and from nearly 80 countries, moves 58,000 tons of cargo per day to and from more than 220 countries, creates jobs for almost 750,000 employees, and powers 28,000 flights across the globe, according to Airline for America [1].

Like any other region in the world, the US was severely affected by the Covid-19 pandemic and airlines are one of the most affected industries. According to Department of Transportation (DOT) statistics, passenger traffic declined 90% in April 2020 as compared to the previous year [2].

Due to a rise in demand for travels in 2021, the industry had a short rebound and later was slowed down because of the delta variant. The cargo airlines have still retained almost the same output and reached the all-time-high in 2021 as seen in Figure 1, which shows the change in passenger and cargo airline percentage between 2019 and 2022 of

[1] Patara Charnvanichborikarn, Department of Industrial Engineering Faculty of Engineering, Chulalongkorn University; 6230221321@student.chula.ac.th

U.S. passengers and cargo airlines. The blue line represents passenger traffic and the brown line represent cargo traffic. However, the overall airline industry could not secure positive profits after the pandemic as compared to other industries in the US as seen in Figure 2.

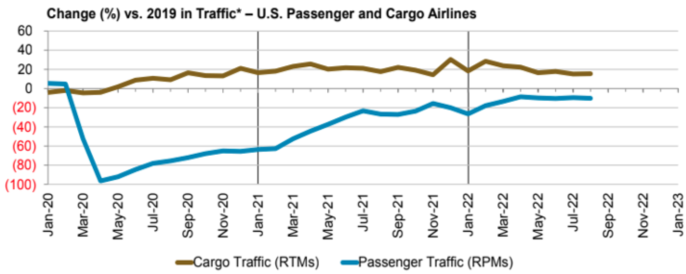

Figure 1. US Cargo Traffic and Passenger Traffic [3]

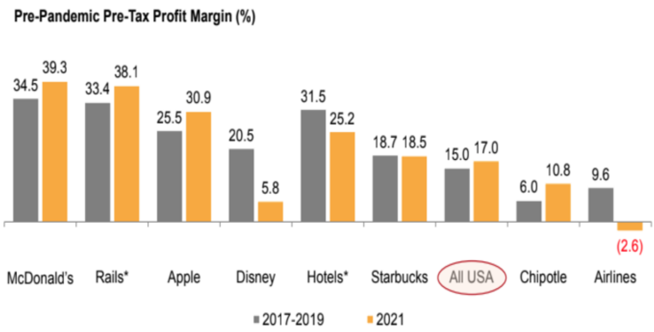

Figure 2. US Industries, 2021 Profitability [3]

The US airlines is one of the largest industries in the world. It is highly competitive and has been facing economic situations as well as financial distress for the past years. Not only that it has to handle high traffic each day, but also that it has to minimize the cost in order to win against other competitors. As a result, airlines have implemented policy changes or new strategies to boost their efficiency in the line of work. Some of them have initiated new marketing campaigns and/or products and services to enter new market segments and maximize their profitability. However, delays and cancellations are found to be an important issue that disrupt the improvement of the airlines in several areas ranging from cost-reduction to customer satisfaction.

1.2. Weather Delay

The aircraft delay causes setbacks and increases the unnecessary load to the air traffic controller since the aircraft is already airborne and eventually the plane needs to get back on the ground. Each airplane time in the air is limited by how much fuel it has. The more fuel, the more time, but it also increases the unnecessary weight and the cost of operating the flight. Each flight carries an optimal amount of fuel with necessary spare. If the plane runs out of fuel, they could potentially cause more problems and become an emergency aircraft itself.

According to the National Airspace System (NAS), weather is the largest cause of aircraft delay. As referred to Fig. 3, it accounts for 75.48% of the delays of greater than 15 minutes over the six years from June 2017 to May 2022 [4].

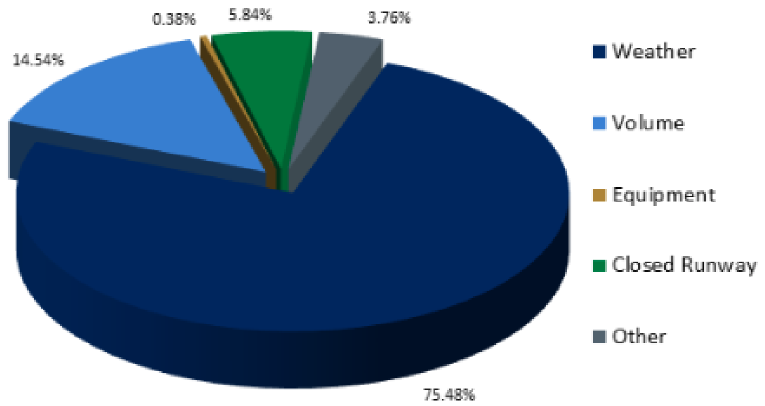

Figure 3. Different causes of aircraft delays in the National Airspace System from June 2017 to May 2022

As the weather is the most unpredictable component of a flight, if there is a slight chance of predictability in what the weather is going to be, it will provide the logistical issue that could be used to prevent further delays. Therefore, aircraft arrival prediction could potentially be a solution to improve the efficiency of the operation and the effectiveness of the contingency plans to respond to such events.

The rest of the paper is organized as follows: The next section summarizes literature review on the evaluation of the impact of delayed departure flights. Section 3 introduces different models used to predict aircraft delays. Section 4 presents the results and discussion about the comparison of the models and identifies the most suitable model. Section 5 concludes the study and provides recommendation.

2. Literature Review

Different airlines have their ways of dealing with delayed aircrafts. Airlines have been overestimating their flight time to make sure that they are on time. This is the current solution that most airlines are doing to compensate for the lost time. The airlines will want to have a flight delay prediction model to improve their schedule, build trust and increase efficiency.

A traditional way is to simply focus on flight time equal to distance divided by velocity. After the time is calculated then the delay factor is added onto the total time. If the aircraft is delayed often then the scheduler will then make an adjustment to the schedule to accommodate the longer commute time. Nevertheless, the airline does not have control over the weather and their effect on aircraft delay. Therefore, a lot of weather data and weather prediction models are available.

Aircraft scheduling buffer has been a widely used method to compensate for aircraft flight delay. It has been used to add more time to the schedule for multiple types of delay. Brueckner et al. utilized the method of airline scheduling buffer choice and studied the shocks influencing flight times [5].

Brueckner et al. provided an insight on the theoretical model to analyze the airline's choice of buffering method, and also suggested that a mitigation of delay propagation could be recovered entirely by the ground buffer and the second was flight buffer [6].

Hajko and Badánik suggested both benefits and the drawback of the airline scheduling buffer, while using artificial data instead of actual flight data. These specific components caused delays in passenger and baggage, cargo and mail, aircraft and ramp handling, technical and aircraft equipment, damage to aircraft & automated equipment failure, flight operations and crewing and other airline related causes. As the authors mentioned, weather is not the reason for every flight delay, but when they do the delay is significant. [7]

Another method that is not widely used is to upgrade the aviation infrastructure. Zou and Hansen suggested this method and discovered that there is a balance in the complicated set of adjustment between passenger demand, air fare, flight frequency, aircraft size, and flight delays which will lead to an equilibrium. Currently there are so many aircraft that can land at the same time, while building another runway or an airport will increase the capacity of the aviation system [8].

Flight delays have been a significant factor at each airport. It has a huge impact on the economic cost and consequences. Federal aviation authorities and a lot of research have been conducted to measure aviation delay and their economic impact. In additions, reports of how much weather affects the environmental cost are widely available [9].

It is significant that each delay must be stopped before it starts to propagate and affect other flights. It is apparent that each aircraft delay causes further delays in the schedule. Schedule buffer has its downside as well. Buffer makes the flight schedule and turnaround time longer than necessary. It also reduces the utilization of the aircraft and drives the operating cost higher. Further study is needed in order to categorize what is important to cause the aircraft delay and what are the necessary information to be included into the artificial network. The necessary information will increase the artificial network performance [10].

Since the world is going to more personalize traveling, smaller aircrafts are preferred. More and more smaller body aircrafts are in demand. On the other hand, the aviation infrastructure has limited resources, and larger aircraft is preferred. This paper would not focus on aircraft maintenance, which is mostly predictable and prepared for as well as the load on the aviation infrastructure that could be calculated. Therefore, this paper would not focus on increasing investment to the business of aviation but would focus on weather delay predictions.

At the end of the day, it all comes down to how much money is there to generate and what is the breakeven point for the airline. The industry is open to finding a possible solution and cure to each specific type of delay. The most accurate model will save a lot of money for the airline.

3. Methodology

The purpose of this section is to introduce the methods used in this paper. First, the data is collected and cleaned to isolate out the relevant factors for aircraft delay. Second, the machine learning based models are built to predict the aircraft delays. Third, the model that predicts the aircraft delays with the highest accuracy and F-1 score is selected.

3.1. Data Collection and Cleaning

Provided by Kaggle, the data consists of multiple files of airline, weather, airport, and employment information. The data that was collected are weather, passenger, aircraft, coords (coordinate for airports), names of carrier, and employee. The weather data is provided by the National Centers for Environmental Information (NOAA). The flight data is provided by the Bureau of Transportation statistics. The data is merged together using date, time, and the link to all of the information. For duplicate data in airport coordinates, names of the carrier, aircraft tail number, and origin airport ID are dropped. The data is collected throughout the years based on the time collected.

There are a total of 77,350 lines for weather data. The weather information includes station name, date, average wind speed, peak gust time, precipitation, average temperature, maximum temperature, minimum temperature, fog, heavy fog, thunder, ice pellets, hail, glaze, dust, smoke, blowing or drifting snow, tornado, high winds, blowing spray, mist, drizzle, freezing drizzle, rain, freezing rain, snow, unknown precipitation, ground fog, and ice fog. Most of the weather information is very specific and not a number or NaN. Every time there is an occurrence of a specific weather phenomenon, it is recorded as 1. Therefore, with the current set of data, NaN is replaced with 0. For general missing weather data such as minimum temperature, average temperature, and average daily wind speed, data is replaced with median for temperature related data and mean.

Next, we define a function which performs feature engineering, data merges and cleanup, using one month of on-time data at a time, which is from the Bureau of Transportation Services. The parameters include monthly_data, aircraft, airport coordinates, names, weather, passengers, and employees. The function also returns cleaned data of one month of on-time reporting. This step includes starting a timer to track how long the cleaning function takes, cleaning up by dropping rows with no departure time, tail number, or canceled. Then, we list flight segment number for daily flight segments by tracking tail number, list the number of concurrent flights at the airport in the time block, and combine the number of seats with the main frame on tail number. The missing aircraft will be filled with the average. Next, we change the data type for the number of seats to reduce the memory usage, merge to get the proper carrier name, add monthly flight statistics for carrier, airport information, and airport flight per month, and add monthly passenger statistics for carriers and airports. After that, we merge the employee, flight attendants per passenger, and ground service per passenger to the data set. For the plane age data, it is calculated by subtracting the current year with the manufacture year of the aircraft (the current year - the manufacture year of the aircraft). The airport coordinates of latitude and longitude of the departing airport are merged and added to the data. The airport that has flight traffic of less than 10th percentile is going to be dropped due to lacking the amount of traffic when compared to other airports.

Meaning that airports that have less than 1,100 flights per month are dropped as these airports are capable of handling manageable traffic volume.

For the selected airports, the weather data is added. After that, the flight data is merged and the columns that are not going to be used are dropped. The dropped columns include origin, destination, course departure time, departure delay new, course arrival time, arrival time, canceled, cancellation_code, course elapsed time, distance, carrier delay, weather delay, nas delay, security delay, late aircraft delay, arrival delay new, arrival time block, actual elapsed time, destination airport ID, destination city name, operator carrier flight number, operator unique carrier, airline ID, date, day of month, tail number, departure time, origin city name, passenger handling, origin airport ID, manufacture year, departures performed for year, and passengers enplaned for year. Next, the data types for all the various fields are cleaned up to reduce memory usage. Month and day of the week are defined as 'object'. For delay, distance group and segment number are defined as 'int8'. Airport flights per month, airline flight month and airline airport flight per month are defined as 'int64'. Lastly, the plane age is defined as type 'int32'. The timer is stopped, and elapsed time is calculated. The train and test sets are generated. The train set is split into subsets and the validation set is generated.

After combining all the data into one file, there are some data columns that are text and not numbers. The method chosen to transform these data is OrdinalEncoder although there are also OnehotEncoder and LabelEncoder. OnehotEncoder creates a binary column for each category and returns an array of numbers. The LabelEncoder transforms the data from text providing values between 0 and n - 1. The OrdinalEncoder will find the unique values per feature and transform the data.

3.2. Prediction Models

3.2.1. Random Forest

For the initial analysis from reducing the non-contributing variables of the flight delay, random forest is used as a supervised learning algorithm method, which can be used for classification and regression. The trees are the subsection of the forest. For a forest with a lot of trees, it becomes richer. The random forest creates a decision tree on randomly selected data. This function gets predictions from each tree and produces the best solution. Also, it produces a good indicator of the importance of the data. The advantage of random forest is that it takes all of the information and averages out the prediction, which cancels out the bias.

The random forest classifier available parameters are n_estimators (list of tree classifiers) of 10, 50, 100, 150, and 200. The max features (the number of features to consider when looking for the best split) are auto, square root, and log2. The max depths are 12, 15, 20, 30, 40, or none (let the leaves expanded until the process is complete or until it is lower than min_sample_split). For criterion, they are gini and entropy. GridSearchCV was performed to find the combination for the parameters, which were the n_estimator between 10 and 200, the max_ depth of 12 and 40, and the min_samples_leaf of 2. The combination that GridSearchCV selected were n_estimator of 200, max_depth of 40, and min_samples_leaf of 2. The other parameters that were used are: criterion = gini, min_sample_split = 2, min_weight_fraction_leaf = 0, max_feature = Sqrt, max_leaf nodes = None, min_impurity_decrease = 0, bootstrap =

True, oob_score = False, n_jobs = None, random_state = None, verbose = 0, warm_start = False, class_weight = None, ccp_alpha = 0, and max_sample = None.

Eq. (1) presents the classification outcome when values of 0,1…K-1 for m be the proportion of class k observations in node m. If m is a terminal node, predict_proba for this region is set to p_{mk}. Common measures of impurity are with Eq. (2). The Entropy uses the Shanon Entropy as a tree node to split different criterion, which is also equal to minimizing the log loss. This is shown in Eq. (3).

$$p_{mk} = \frac{1}{n_m} \sum_{y \epsilon Q_m} I(y = k) \tag{1}$$

Gini Impurity Equation:

$$H(Q_m) = \sum_k p_{mk}(1-p_{mk}) \tag{2}$$

Log Loss or Entropy Equation:

$$H(Q_m) = -\sum_k p_{mk} \log(p_{mk}) \tag{3}$$

3.2.2 Adaboost

Adaboost is a method that uses the complete train dataset to train. It is a sequential process. Each of the following models tries to correct the errors of the previous model. The model tries to correct the errors so, for the next iteration, it is better. Each model is trained on the same dataset, but each of the data samples is assigned a different weighting factor in the previous model's success. It is the learning process in sequence. The weighing factor is then reassigned in every iteration to make a better classifier than the previous iteration. The process begins when a subset is selected from the original dataset. Then all training examples are assigned the same weight. A base model is then trained on this subset. The final model of this subset will be used to make predictions on all data. The errors are calculated using the actual values and the predicted values.

For the adaboost method, GridSearchCV was performed to find the combination for the parameters, which were the n_estimator of 50, 100, and 150, and the learning_rate of 0.1, 0.25, 0.5, and 1. The combination that GridSearchCV selected were learning rate of 1 and n_estimator of 150. The algorithm used is "SAMME.R". The SAMME.R algorithm typically converges faster than SAMME and achieves a lower test error with fewer iterations. The random_state is none due to the controlling of randomness of the result. The base estimator is none because it is a decision tree classifier with a max_depth equal to 1.

4. Results and Discussion

From the collected data, the methods that are chosen to construct the prediction models are random forest and adaboost. The preferred prediction method in this case between random forest and adaboost is random forest, which is selected according to the accuracy and F-1 score of both models in Fig. 4 and Fig. 5. There are possible room to improve the model for a better F-1 score. Different parameter effects the result of the prediction model. The adjustment can also be made to the hyperparameters or features of any classifiers to better optimize the model. Furthermore, the data collected lacks real time weather data. This could be incorporated into the data set for a better prediction.

	precision	recall	f1-score	support
0	0.93	0.99	0.96	737142
1	0.94	0.67	0.78	171321
accuracy			0.93	908463
macro avg	0.93	0.83	0.87	908463
weighted avg	0.93	0.93	0.93	908463

Figure 4. Classification Report Random Forest Results

	precision	recall	f1-score	support
0	0.81	1.00	0.90	3684082
1	0.56	0.01	0.02	858232
accuracy			0.81	4542314
macro avg	0.69	0.50	0.46	4542314
weighted avg	0.76	0.81	0.73	4542314

Figure 5. Classification Report Adaboost Results

The prediction accuracy result for random forest is 92.98 percent and it is 81.14 percent for adaboost. These numbers are rounded off and then displayed in Fig. 4 and Fig. 5. Both models have high accuracy but random forest is higher by 11.84 percent. F-1 score is composed of precision and recall. Once it is calculated, it is shown on Fig. 4 and Fig. 5 for each model. The higher F-1 score is considered to be a better model overall. When comparing the F-1 score, macro average for the adaboost and random forest is used. Adaboost macro average is lower than the random forest macro average by 0.41. This result suggests that the random forest should be the chosen method. However, there is room for improvement and fine tuning for the model. For example, the n_estimator can be adjusted, feature adjustment is necessary to improve the accuracy of the model. Furthermore, there will also be a point where different combination of feature adjustments can reduce the accuracy as well.

5. Conclusion and Recommendations

In conclusion, the aircraft delay has cost the airline industry time and money. Once the first aircraft is delayed, it propagates down the whole schedule. It triggers a reaction of which it is hard to recover from. The airlines have initiated the idea of including extra time in their schedule, or in another term called schedule buffer. Several ways are used to calculate buffer and each airline chooses a different one. The most common method that most airlines use is by using the average time and adding a little extra buffer.

Weather is one of the most contributors of aircraft delays. In order to predict the optimum schedule buffer for the airline the prediction, a model for aircraft weather delay has to be created. Currently, the weather data is the data that is collected for the entire day. With the daily weather data, it provides a gap for future big picture predictions. To help with the prediction in the future, more available data and real time weather sensors can predict the flight delays with higher accuracy. Additional information can be collected from the winds aloft and the wind at local airports. It can be updated hourly or as necessary. These results are the crucial decision factors if the pilot is going to keep the plane airborne or make the critical decision whether the aircraft is going to be delayed due to the weather.

NOAA provides the current weather data for this model. There is aviation data as well such as Automatic Terminal Information Service (ATIS), Automated Surface Observing Systems (ASOS), Automated Weather Observing System (AWOS), Meteorological Terminal Air Report (METAR), and Terminal Area Forecast (TAF). These are some of the systems that have been put in place for the aviation community in the United States in order to monitor the surface weather. AWOS measures information such as wind speed, wind gust, and wind direction. These are limiting factors to which an airplane could land at the airport or not. The information is broadcasted over the internet as well as on landline. Some of the observations are also produced by humans. Winds aloft is wind speed at altitude collected from National Centers for Environmental Prediction (NCEP). Currently, the winds aloft data are not covering all of the flight levels. The winds aloft data are available on the NOAA website as well.

Reference

[1] Airline for America. "U.S. Passenger Carrier Delay Costs." DATA & STATISTICS U.S. Passenger Carrier Delay Costs, Airlines For America, 12 July 2022, https://www.airlines.org/dataset/u-s-passenger-carrier-delay-costs/.

[2] Office, U.S. Government Accountability. "Covid-19 Pandemic: Observations on the Ongoing Recovery of the Aviation Industry." COVID-19 Pandemic: Observations on the Ongoing Recovery of the Aviation Industry | U.S. GAO, 9 Nov. 2021, https://www.gao.gov/products/gao-22-104429.

[3] Airline for America. "U.S. Passenger Carrier Delay Costs." DATA & STATISTICS U.S. Passenger Carrier Delay Costs, Airlines For America, 12 July 2022, https://www.airlines.org/dataset/u-s-passenger-carrier-delay-costs/.

[4] FAA. "FAQ: Weather Delay." FAQ: Weather Delay | Federal Aviation Administration, FAA, https://www.faa.gov/nextgen/programs/weather/ faq.

[5] J. K. Brueckner, A. I. Czerny and A. A. Gaggero "Airline Schedule Buffers and Flight Delays: A Discrete Model." Economics of Transportation, Elsevier, 20 July 2021, https://doi.org/10.1016/j.ecotra.2021.100218.

[6] J. K. Brueckner, A. I. Czerny and A. A. Gaggero "Airline mitigation of propagated delays via schedule buffers: Theory and empirics" Transp. Res. E, 105 (2021), Article 102333

[7] J. Hajko and B. Badánik "Airline on-time performance management." Transportation Research, Procedia. 4 Dec. 2020, https://doi.org/10.1016/j.trpro.2020.11.011.

[8] B. Zou and M. Hansen. "Flight Delays, Capacity Investment and Social Welfare under Air Transport Supply-Demand Equilibrium." Transportation Research Part A: Policy and Practice, Pergamon, 23 Mar.2012,https://www.sciencedirect.com/science/article/pii/S096585641200033X

[9] D. M. M. S. Dissanayaka, V. Adikariwattage and H. R. Pasindu "Evaluation of Emissions from Delayed Departure Flights at Bandaranaike International Airport (BIA)." Evaluation of Emissions from Delayed Departure Flights at Bandaranaike International Airport (BIA) | Atlantis Press, Atlantis Press, 1 Oct. 2019, https://doi.org/10.2991/apte-18.2019.26.

[10] Pankaj Kumar and Ajai Kumar Tiwari, "Evaporation Estimation Using Artificial Neural Network," *International Journal of Computer Theory and Engineering* vol. 4, no. 1, pp. 51-53, 2012.