Predicting Airport Pushback Times and Taxi Times Using ML Algorithms

Jihan Song, Lisa Ko, Audrey Dahlkemper, Nicola Rowe

I320D Applied Machine Learning with Python

Intro

Aviation has become increasingly accessible over the past several decades due to advances in the technology and flight fields. As a result, air travel has exponentially increased. The Federal Aviation Administration (FAA) estimates that it services more than 45,000 flights a day in the US (FAA, n.d.). However, as the demand for flights increases, the delays and complications in air traffic increase as well. The air transportation system is extremely vast and complex. A single delay often alters the path of other flights and exacerbates issues in current airline operations systems. Our analysis focuses specifically on a project conducted by NASA on the "Prediction of Pushback Times and Ramp Taxi Times for Departures at Charlotte Airport" (Lee et al., 2019). The researchers created six different machine learning algorithms for pushback time prediction: Linear Regression, Support Vector Regression, Lasso Linear Regression, K-Nearest Neighbors, Random Forest, and Neural Networks. Our project goal is to determine the effectiveness of these different algorithms and factors for airport pushback times and taxi times then create a new model based on our evaluations.

Examination of Data

The project is split into four sections. First, a detailed time analysis of taxi time and the pushback process is performed using ATD-2 system data. The Airspace Technology Demonstration 2 (ATD-2) integrates arrival, departure, and surface operations. Next, the researchers formulated a decision tree representative of the pushback process time. Then, a simple prediction model was created that focuses specifically on transit time. Last, machine learning models are applied to evaluate the accuracy of the prediction model. Taxi-Out time was selected as the primary feature because it is used to predict the earliest possible takeoff time. It is also commonly recognized as the factor that overarchingly contributes to the majority of delays. The variable is defined as the "total amount of time a charter jet spends moving on the ground" (Stratos Jet Charters, n.d.). Within the context of this ML project, taxi-out time was split into two parts - airport movement area and ramp area. Charlotte Airport was chosen because it frequently incurs long taxi times. In understanding what causes the extended ground time, future researchers can implement more effective strategies that decrease pushback time and increase economic profit. In total 67 features were defined and utilized in the machine learning algorithms. Selected factors include aircraft type, carrier, actual pushback time of a day in hours, gate conflict, and restriction of Expected Departure Clearance Time (EDCT) and Approval Request (APREQ). Two-thirds of the data was used as a training dataset while one-third was used as a test dataset. Data for the models were collected over a month—from August 1, 2018, to August 31, 2018.

Pushback times were found to be mostly dependent on gate location and aircraft type. Out of the two prediction models, the one with ramp area and aircraft type returned better results. Among the different algorithms there was no "significant difference" (Lee et al., 8) in the prediction accuracy. The mean and median values of the prediction errors are almost zero or less than 25 seconds. Moreover, the MAE and RMSE values suggest that the deviations of the predicted pushback times from the actual ones are around 90 seconds regardless of methods.

Shortcomings Of The Data

Flight operations are subject to a myriad of factors, many of which are uncertain. Therefore, creating the perfect predictive model for pushback times is extremely difficult. The optimal strategy would involve creating an extremely complex machine learning algorithm that incorporates intrinsic and extrinsic factors such as the technology utilized and weather conditions. Regarding the data from the research project, there were three main shortcomings: period, airport applicability, and limited extrinsic features.

The data used in the project was collected over one month from August 1, 2018, to August 31, 2018, which may not be representative of the entire year. This limited time frame could potentially affect the generalizability of the findings to other periods. Second, the study only focuses on Charlotte Airport as a result the findings may not be generalizable to other airports. Different airports may have different factors affecting pushback times and ramp taxi times, such as different runway configurations, weather patterns, and air traffic volumes. Lastly, the study considers a limited number of extrinsic factors that affect pushback times and ramp taxi times. One of the most influential features is the weather. The impact of weather on pushback times and ramp taxi times was not considered in the study despite the high correlation found in other research on weather and delay time. Weather conditions, such as wind, fog, or snow, can significantly affect airport operations and lead to longer pushback times and ramp taxi times. Incorporating weather data into the analysis could provide more accurate predictions and insights into how weather patterns impact airport operations. Moreover, a range of other factors can affect airport operations such as airline crew availability, baggage handling, and security checks. A more comprehensive analysis of the various factors that contribute to airport delays would provide a better understanding of the challenges faced by airport operators and airlines.

Solutions

1. Weather: To incorporate weather data into the analysis, the researchers could obtain historical weather data for the period under investigation and use it as an additional

feature in the machine learning algorithms. Alternatively, they could use real-time weather data to adjust the predicted pushback times and ramp taxi times in response to changing weather conditions

2. Different datasets: To address the limited scope of the dataset, the researchers could consider using data from multiple airports over longer periods. They could also consider data from different regions, different types of aircraft, and different airlines to capture a more comprehensive understanding of the factors that contribute to pushback times and ramp taxi times
3. Limited scope: To address the limited scope of the study, the researchers could consider including additional factors that could impact pushback times and ramp taxi times, such as airline crew availability, baggage handling, and security checks. They could also consider a broader range of machine learning algorithms to determine which one provides the best predictions.

Creating a new model

The data collection to create a model was one of the most important and tricky parts of the research. Since ATD-2 data is not available to the public, our team decided to find the data that is the most similar and relevant to the original dataset. After research, we found the data from drivendata.org, which contains various information such as GUFI (global unique flight identifier), departure runway actual time (the time that the flight departed from the runway), aircraft type, et cetera. The dataset is only available to competitors under the rule and is not available to the public.

After transforming the data, we faced two major problems. First, the data did not contain information about when and where the ramp was for each aircraft. This information is important because the total taxi out time (total time for an aircraft to be pushed back and depart) consists of three different parts: the original area to the ramp, from the ramp to waiting in the queue, and from the queue to the runway. However, the dataset only contained the time when an aircraft was pushed back (departure stand actual time) and when it departed from the runway. Thus, the original purpose of the research, to create two separate models to estimate (1) pushback time and (2) ramp taxi time, is determined impossible.

Second, the important features were missing. Since having total taxi out time is the label, we had to prioritize having time data and remove NA or NaN values. However, after trimming the data, we were missing important features such as major carrier, aircraft type, or any other features that could be used to estimate total taxi out time data except the weather data. Thus, to sum up, we had 583735 different aircraft total taxi out time and weather data only. The original plan failed. However, as seen in Figure 1, the distribution of total taxi out time has a heavily skewed distribution. This implies that there are lots of occasions in which the flights get delayed during the process. So, instead of creating the model to predict pushback time and ramp taxi time, we decided to build a model that classifies whether the total taxi out time will be outliers or not based on the weather data.

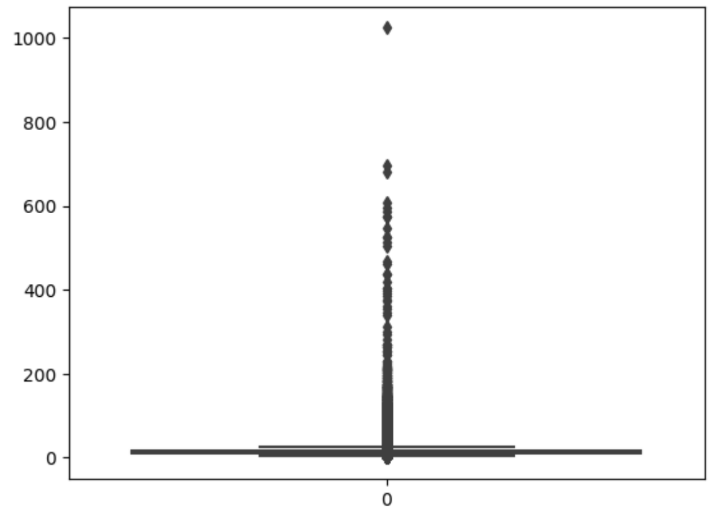

Figure 1

We used an unconventional method for the outliers - 1.5 times interquartile range plus 75%. The mean total taxi out time is 15.9 minutes. The interquartile range is 5.8 minutes. So from 75%, which is 17.73 minutes, the outlier is when the total taxi out time is more than 23.53 minutes. So, in total, amongst 583735 data points, 29557 were outliers—about 5%.

The features were mixed with continuous - temperature, wind direction, wind speed, wind gust, cloud ceiling, and visibility - and categorical - cloud, lightning, and precipitation - data. Since we already lost most of the data from the transformation part, we did not drop any features to have as much data as possible. We chose a histogram gradient boosting classifier for the classifier since it can naturally deal with NaN values and is faster than a normal gradient boosting classifier.

	Predicted_True	Predicted_False	
Actual_True	55375	2856	58231
Actual_False	27	116	143
	55402	2972	58374

The training and testing accuracy of the model was both about 0.95, which is a fairly high accuracy. From a total of 583735, 58374 random aircraft were selected for the test (10%). Among them, 97% were non-outliers while only about 3% were outliers. The model did a good job of predicting the non-outliers. The true positive rate is 95%, and the true negative rate is 81%. Considering that the size of the outlier is fairly small and that weather was the only available data, the true negative rate of 81% is fairly good accuracy. From the ablation test, we found that only dropping “wind gust” and “cloud” increased the accuracy.

The model could be implemented by NASA to detect whether they should be warned of possible delays. Also, with a better and broader amount of dataset, they might be able to build a model that outputs accurate numerical measures in minutes. The current model’s problem is that it only takes the weather data from when it departs from the gate (so when taxi out time starts), so it is not actively reflecting the weather report. If the model is improved, we could build a model that can use real-time data to predict the pushback time and ramp taxi time regarding the weather, which the authors could not accomplish.

Conclusion

The research project by NASA that focuses on predicting pushback times and ramp taxi times for departures at Charlotte Airport using machine learning algorithms. The study used 67 features to evaluate six different algorithms for pushback time prediction. The findings suggest that pushback times were mostly dependent on gate location and aircraft type. However, the data has several shortcomings, including a limited period, airport applicability, and limited extrinsic features such as weather conditions. Incorporating more comprehensive data could provide a better understanding of the factors that contribute to airport delays. To overcome these shortcomings, we suggested using data from multiple airports over longer periods and using historical or real-time weather data to adjust predictions. Overall, the study provides insight into how machine learning can be used to predict pushback times and ramp taxi times, but further research is needed to address the limitations of the data.

We attempted to create a model to estimate pushback time and ramp taxi time for aircraft departing from the ATD-2 site. However, due to the unavailability of the ATD-2 data and missing features in the alternative dataset, we shifted our focus to building a model that classifies whether the total taxi out time will be an outlier or not based on weather data. Our model achieved a high accuracy of about 0.95 in both training and testing, with a true positive rate of 95% and a true negative rate of 81% for outlier detection. While the model's current implementation only considers weather data at the start of taxi out time, further improvements could be made to incorporate real-time weather data and estimate more accurate numerical measures in minutes. Future research could explore additional features and larger datasets to further improve the model's accuracy and applicability.

References

Federal Aviation Administration. (n.d.). *Air traffic by the numbers.* Retrieved April 10, 2023, from https://www.faa.gov/air_traffic/by_the_numbers

Stratos Jet Charters, Inc. (n.d.). *Taxi time.* Retrieved April 10, 2023, from https://www.stratosjets.com/glossary/charter-aircraft-taxi-time/

Lee, H., Coupe, J., & Jung, Y. C. (2019). *Prediction of pushback times and ramp taxi times for departures at Charlotte airport.* AIAA 2019-2933. https://doi.org/10.2514/6.2019-2933