33RD CONGRESS
OF THE INTERNATIONAL COUNCIL
OF THE AERONAUTICAL SCIENCES
STOCKHOLM, SWEDEN, 4-9 SEPTEMBER, 2022

ICAS 2022 SWEDEN

MACHINE LEARNING METHODS ENSURING BOTH PERFORMANCE AND INTERPRETABILITY OF ESTIMATING AIRCRAFT ARRIVAL TIMES

Nobuharu Morikawa[1] & Eri Itoh[1,2]

[1]Faculty of Aerospace Engineering, The University of Tokyo
[2]Air Traffic Management Department, Electronic Navigation Research Institute

Abstract

Estimating the arrival time of aircraft more accurately is one of the most important functions for optimizing the runway operation. For this purpose, this study proposes applying machine learning techniques and aims to clarify the most feasible method which minimizes the estimation errors. In this paper, we applied five supervised learning methods; Elastic Net, Decision Tree Regression, Random Forest Regression, Gradient Boosting Decision Tree, and support vector machine, for estimating aircraft arrival time at Tokyo International Airport. The results showed that the Gradient Boosting achieved the best performance under the given assumptions. Furthermore, Permutation Feature Importance(PFI), Partial Dependence(PD), Individual Conditional Expectation(ICE), and SHapley Additive exPlanations(SHAP) values are introduced to further understand the results estimated using the ML methods. Based on the feature analysis and these indices, we propose a methodology to implement the ML model into the AMAN/DMAN system in the actual air traffic operation

Keywords: supervised machine learning, air traffic management, SHAP

1. Introduction

Although air traffic demand is currently declining in the short term due to the impact of COVID-19, in the long term, it is estimated to increase by 2.2 times in 2039 compared to 2019 [1]. Efficient runway management is required to further improve arrival and departure aircraft traffic management at the airport. In our previous research, we proposed the design requirements for an aircraft arrival management system (AMAN: Arrival Manager) for Tokyo International (Haneda) Airport [2–7]. To further improve the efficiency of runway management, our study indicated that the prediction of arrival time at the runway in the en-route airspace (about 150NM to 200NM away from Haneda Airport) can be shared with the departure management system (DMAN), which has the function of managing the time of aircraft departing from Haneda Airport, thereby contributing to the efficient use of the runway at the airport.

With this background, we have been developing a methodology applying Machine Learning (ML) models to improve the accuracy of arrival time estimation in the integrated AMAN/DMAN system. Our proposing method contributes to predicting the aircraft arrival time approximately 30 minutes before landing and takeoff. The proposing method uses flight information, which allows us to obtain during the actual operation while the arrival aircraft is flying in the en-route airspace. Furthermore, the stakeholders involved in airport operations require us to explain the outputs of the ML models. For ensuring both estimation accuracy and interpretability, we applied elastic net, decision trees, and support vector machines in this study. The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 describes the data used in this study. Section 4 states the problem and reviews each machine learning method, while Section 5 presents our experimental evaluation. Section 6 draws concluding remarks.

2. Related Work

Significant work has gone into applications involving machine learning for air-traffic management, in areas such as air-traffic flow networking and trajectory classification [8–10], delay prediction for air traffic networks [11, 12], dynamic airspace sectorization [13], trajectory prediction in terminal airspace [14] and runway exit prediction for arriving aircraft [15]. These studies have shown scope for machine-learning methods utilizing airtraffic data are effective in extracting and understanding complex airtraffic characteristics, but no results have been published on their application to the operation and design of airtraffic management systems. Accordingly, this study investigates a machine learning method suitable for predicting aircraft arrival times in en-route airspace as part of efforts to contribute to the design and operation of AMAN/DMAN. To predict arrival time more accurately, various machine-learning methods are applied by incorporating factors related to air-traffic management, such as airspace capacity, into the feature set.

3. Track Data of arrival aircrafts at Haneda Airport and Feature Extraction

3.1 Applying a distance-based Model

In this research, we used flight plans and track data for 39 days between September 2019 and February 2020 during northerly wind operations arriving at Haneda Airport from the southwest. Figure 1 shows the track data, along with concentric circles with radii varying every 10 NM from 10 to 200 NM starting from Haneda Airport and the outer edge of the terminal airspace. Arrival flows from the southwest are indicated by red lines.

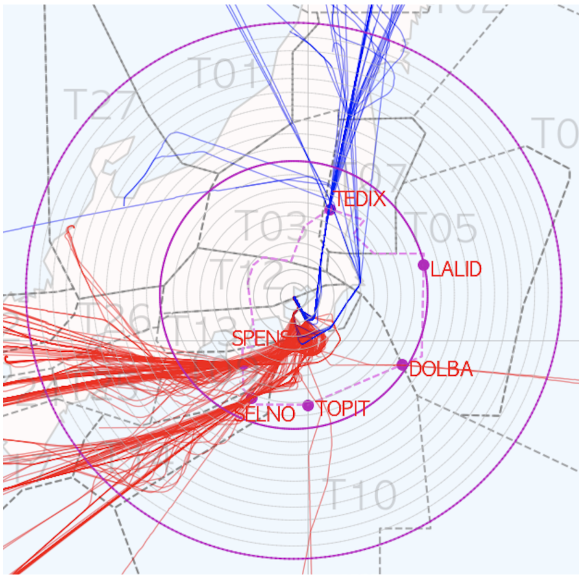

Figure 1 – Concentric circles at the center of Haneda Airport and arrival traffic flow model (Capital letters indicate waypoints to be transferred to the terminal airspace.)

Here, we define 20 airspaces $i \in \{1,2,\ldots,20\}$ surrounded by concentric circles. Airspace $i = 1$ is defined as an airspace surrounded by concentric circles of radius 10 and 20 NM, while airspace $i = 2$ is defined as one surrounded by concentric circles of radius 20 and 30 NM and so on.

3.2 Feature Extraction

3.2.1 Numerical Variables

For each airspace i defined in the previous section, the flight time taken by aircraft arriving at Haneda Airport in each airspace, as well as the ground speed, altitude, separation, namely the time interval from the preceding aircraft and heading, measured when entering each area of airspace, were extracted and used as numerical variables. The separation represents the degree of congestion in the airspace. Table 1 summarizes the numerical variables.

Table 1 – Numerical variables

Feature	Description
XXXtoYYYFT	Flight time taken from XXX NM to YYY NM
XXXGS	Ground speed at XXX NM
XXXAL	Altitude at XXX NM
XXXSEP	Separation at XXX NM
XXXANG	Heading at XXX NM(0-360 degrees)

3.2.2 Categorical Variables

Categorical variables are defined as shown in Table 2 for factors that may affect air traffic. The selection of categorical variables is based partially on literature [12], and we use the airline, departure airport, RECAT(Wake Turbulence Re-categorization [16]), transit gate (entry point into terminal airspace) and runway arrival time.

Table 2 – Categorical variables

Classification	Feature	Description
Airline	airline 1	Airline 1
	airline 2	Airline 2
	airline 3	Airline 3
	airline 4	Airline 4
	airline 5	Airline 5
	other_airline	Other Airlines
Departure Airport	RJOOBB	Osaka International Airport or Kansai International Airport
	RJFF	Fukuoka Airport
	ROAH	Naha Airport
	RJCC	New Chitose Airport
	RJGG	Chubu Centrair International Airport
	OTHER_JAPAN	Other Airports in Japan
	ASIA	Airports in Asia
	Europe	Airports in Europe
	Hawaii	Airports in Hawaii
	N_America	Airports in North America
	South	Sydney International Airport or Auckland International Airport
RECAT	WV_A	Class A in RECAT
	WV_B	Class B in RECAT
	WV_C	Class C in RECAT
	WV_D	Class D in RECAT
	WV_E	Class E in RECAT
	WV_F	Class F in RECAT
	WV_G	Class G in RECAT
Pass-through Gate	AKSEL	AKSEL
	XAC	XAC
	AROSA	AROSA
	HME	HME
	155E	155E
	STONE	STONE
Arrival Time	XXX	Within the hour of XXX

RECAT is classified into seven categories (A, B, C..., G) based on the aircraft's maximum takeoff weight and wingspan. WV_A is the aircraft with the largest takeoff weight and wingspan, followed by smaller aircraft in alphabetical order. For example, B777 is classified as WV_B, and A320 or B730 is classified as WV_D.

The transit gates are defined by the names of the waypoints via which arriving aircraft next pass after traversing the six waypoints on the outer edge of the terminal airspace shown in Figure 1. The main air traffic arriving from the southwest enters the terminal airspace through SPENS and SELNO, then traverses XAC and AKSEL, respectively.

4. PROBLEM STATEMENT AND MODEL REVIEWS

We use features introduced in the previous section and target a model that predicts the flight time taken from 200 to 80 NM when arrival aircraft pass 200 and 150 NM respectively. In response, we build several models, which are then ranked by performance.

4.1 Elastic Net

The elastic net [17] is a type of multiple regression analysis that optimally exploits ridge and lasso regression. N samples and p features $x_1, \ldots, x_p$, and coefficients $\hat{\beta} = (\hat{\beta}_0, \ldots, \hat{\beta}_p)$ are used to predict the target using Equation1.

$$\hat{y} = \hat{\beta}_0 + x_1\hat{\beta}_1 + \cdots + x_p\hat{\beta}_p \tag{1}$$

The objective function is set as in Equation2 and parameter $\hat{\beta}$ is found to minimize it.

$$L(\lambda_1, \lambda_2, \boldsymbol{\beta}) = |\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}|^2 + \lambda_1|\boldsymbol{\beta}|_1 + \lambda_2|\boldsymbol{\beta}|^2 \tag{2}$$

The elastic net is characterized by variable selection and grouping effects and given its tendency to estimate sparse solutions, over-fitting (over-learning) is unlikely to occur, making it a highly interpretable method. The disadvantage is that as a type of linear regression, it is less accurate for regression problems with strong nonlinearities.

4.2 Decision Tree

Decision trees [18] are widely used models for classification and regression. The former is called a classification tree and the latter a regression tree. In decision trees, classification and regression are performed using a hierarchical tree structure; comprising questions that can be answered with Yes/No.

4.2.1 Regression Tree

The advantages of regression trees are their high readability due to a tree structure and robustness against outliers. Conversely, downsides include instability, since the learning result is prone to change significantly in response to small changes in training data and overfitting can occur unless appropriate pruning is performed. Accordingly, ensemble methods like Random Forest and the Gradient-Boosting Decision Tree are more effective.

4.2.2 Random Forest

The Random Forest [19] involves collecting several slightly different decision trees and averaging them. This is based on the concept that although individual decision trees can allow relatively accurate predictions, they overfit some of the data. If numerous decision trees are created, each overfitting in a different direction, overfitting can be suppressed without pruning by averaging the results.

4.2.3 Gradient-Boosting Decision Tree(GBDT)

Decision trees are created in sequence, with the next decision tree correcting the errors of the previous one in GBDT [20]. The GBDT uses very shallow decision trees around 1 to 5 in depth, which allows the model to occupy less memory and accelerate the computation. The process of adding decision trees one at a time and gradually making ever-diminishing modifications is key to steadily reducing the rate of error in the overall data.

4.2.4 Support Vector Machine(SVM)

SVM can handle non-linear data by margin-maximization and kernel methods. The advantage of SVM is its positive performance, even when the dimension of the data increases and it has fewer parameters to optimize. The disadvantage is that the more training data involved, the more computationally intensive it becomes.

5. Results

5.1 Prediction Error when using the Haneda-centered Model

Table 3 shows the RMSE values when predicting flight time from 200 to 80 NM for each algorithm at 200 and 150 NM from Haneda Airport, along with the RMSE when using the average flight time from 200 NM to 80 NM as the baseline. Table 3 explains that GBDT is the most accurate when it comes to predicting flight time. Nevertheless, there is a relative error of about 10% in GBDT, since the average flight time from 200 to 80 NM is 921 seconds.

Table 3 – RMSE values when Haneda-centered distance-based model is applied.

Methods	Airspace of the features used	RMSE [s]
Elastic Net	$i = 15$	88
	$i = 20$	102
Decision Tree	$i = 15$	93
	$i = 20$	127
Random Forest	$i = 15$	85
	$i = 20$	107
GBDT	$i = 15$	82
	$i = 20$	94
SVM	$i = 15$	87
	$i = 20$	97
Baseline	-	134

5.2 Improving the distance-based Model

Next, we improved the distance-based model by defining an airspace $j \in \{1, 2, \ldots, 20\}$ surrounded by concentric circles with Ohshima as the starting point and verified the accuracy of predicting the flight time at the point of entry into the terminal airspace. Figure 2 shows the concentric circles centered on Ohshima and the traffic flow arriving at Haneda Airport.

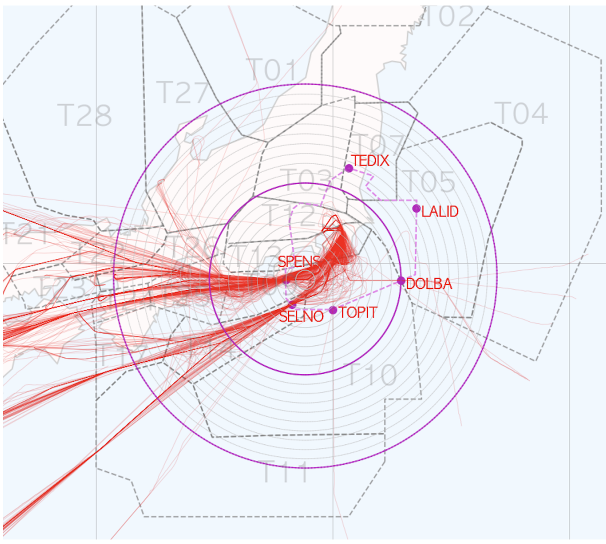

Figure 2 – Map of Japan and concentric circles centered on Ohshima are drawn with flight tracks. The red tracks show the south-west traffic flow

As shown in Figure 1 and 2, traffic arriving at Haneda Airport from the southwest is split into two flows, traversing the two gates, SPENS and SELNO. In the Haneda-centered model, the difference in flight time is attributable to the location of the transit gates for these two traffic flows. By using the Ohshima-centered model, we hypothesized that the two traffic flows could be treated more equally and accurately. However, since 80-150-200 NM in the Haneda-centered model corresponds to 30-

120-180 NM in the Ohshima-centered model, we compare the results using corresponding airspace features.

5.3 Prediction Error when using the Ohshima-centered Model

Table 4 shows the RMSE values when predicting the flight time from 180 to 30 NM for each algorithm at 180 and 120 NM from Ohshima, along with the RMSE baseline. Table 4 explains that the elastic net improved accuracy by about 12-15%. Since other methods also helped boost accuracy significantly, it can be analyzed that the two main traffic flows can be treated equally by improving the distance-based model and extracting features at the center of Ohshima, thereby reducing the dimensionality of categorical variables, given scope to eliminate the effect of the two pass-through gates (SPENS and SELNO).

Table 4 – RMSE values when the Ohshima-centered distance-based model is applied.

Methods	Airspace of the features used	RMSE [s]
Elastic Net	$j=12$	73
	$j=18$	90
Decision Tree	$j=12$	75
	$j=18$	123
Random Forest	$j=12$	65
	$j=18$	75
GBDT	$j=12$	63
	$j=18$	77
SVM	$j=12$	64
	$j=18$	78
Baseline	-	134

5.4 Interpretability

5.4.1 Permutation Feature Importance(PFI)

PFI involves calculating the importance of a feature in a prediction by examining the extent to which prediction error increases when the value of a feature is made unavailable. PFI can be easily calculated, no matter how complex the model, like Random Forest. Specifically, a feature is first rendered unusable by shuffling its columns. The importance of the feature is then determined by calculating the incremental prediction error. This operation is performed several times and the average is taken as the importance of the feature.

Figure 3 shows the PFI plot when predicting the flight time taken from 200 to 80 NM at 200 NM in a Random Forest by the Haneda-centered model.
The Numerical variables are at the top of the list, with ground speed at 200 NM (200GS) the key parameter. Variables such as the flight time taken from 210 to 200NM (210to200FT) and heading at 200 NM (200ANG) are relatively high compared to other numerical values. The categorical variables such as the departure airport and arrival time zone impact on the prediction, despite being relatively small compared to numerical ones.
Conversely, Figure 4 shows the PFI plot when predicting the flight time taken from 180 to 30 NM at 180 NM in a Random Forest by the Ohshima-centered model.
Unlike the Haneda-centered model, the ground speed at 180 NM (180GS) prevails and other values are smaller. As mentioned earlier, this is because the Ohshima-centered model allows two separate traffic flows to be treated equally, thus reducing complexity.

5.4.2 Partial Dependence(PD)

PD is a powerful method to generally determine the relationship between features and predictions. It is an interpretation method that fixes other features, moving only those specified, then averaging and visualizing the predictions of each instance.

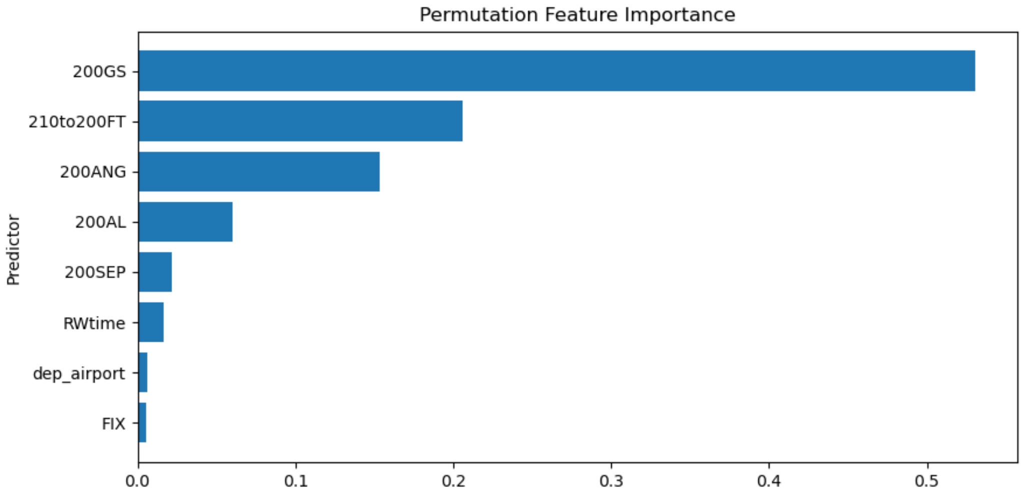

Figure 3 – Permutation Feature Importance Plot when predicting the flight time taken from 200 NM to 80 NM at 200 NM in a Random Forest by the Haneda-centered model

Figure 5 represents the PD values plot of the numerical variables and Figure 6 the PD values plot of the categorical variables.
Figure 5 shows that the flight time taken from 190 to 180 NM and that from 180 to 30 NM tend to correlate negatively. Conversely, a ground speed of 180 NM or separation at 180 NM tends to correlate negatively with the flight time taken from 180 to 30 NM.
Figure 6 shows a tendency toward shorter flight times between 4 and 8 o'clock and longer from about 19:00. The PD values also differ depending on the departure airport, with flights departing from Osaka(dep_airport=0) usually taking longer.

5.4.3 Individual Conditional Expectation(ICE)

PD is a good way to generally determine the relationship between features and predictions, but taking the average means the effect of the interaction can no longer be considered. Accordingly, ICE is a method used to check the relationship between features and model predictions for each instance without taking an average.
Figure 7 represents the ICE values plot of the numerical variables and Figure 8 the ICE values plot of the categorical variables.
One thing the PD did not find is that Figure 7 shows declining flight time as the heading approaches 90 degrees, in an eastward direction.

5.4.4 SHapley Additive exPlanations(SHAP)

SHAP, like ICE, renders the predicted value of each instance interpretable. Specifically, SHAP decomposes the difference between the "predicted value for an instance" and the "average predicted value" into the contribution of each feature. $\boldsymbol{X}$ is the feature and $\hat{f}(\boldsymbol{X})$ is the learned machine learning model. Specifically $x_i = (x_{i,1}, \ldots, x_{i,J})$ is the feature of an instance i, and the prediction value of the instance i becomes $\hat{f}(x_i)$. In SHAP, we decompose the difference between the estimation value $\hat{f}(x_i)$ and the average value of the prediction $E[\hat{f}(\boldsymbol{X})]$ into the contribution of each feature. The contribution of the feature $x_{i,j}$ of the instance i is defined as $\phi_{i,j}$, and the difference from the expected value is decomposed into the sum of the feature contribution as Formula 3 represents.

$$\hat{f}(x_i) - E[\hat{f}(\boldsymbol{X})] = \sum_{j=1}^{J} \phi_{i,j} \tag{3}$$

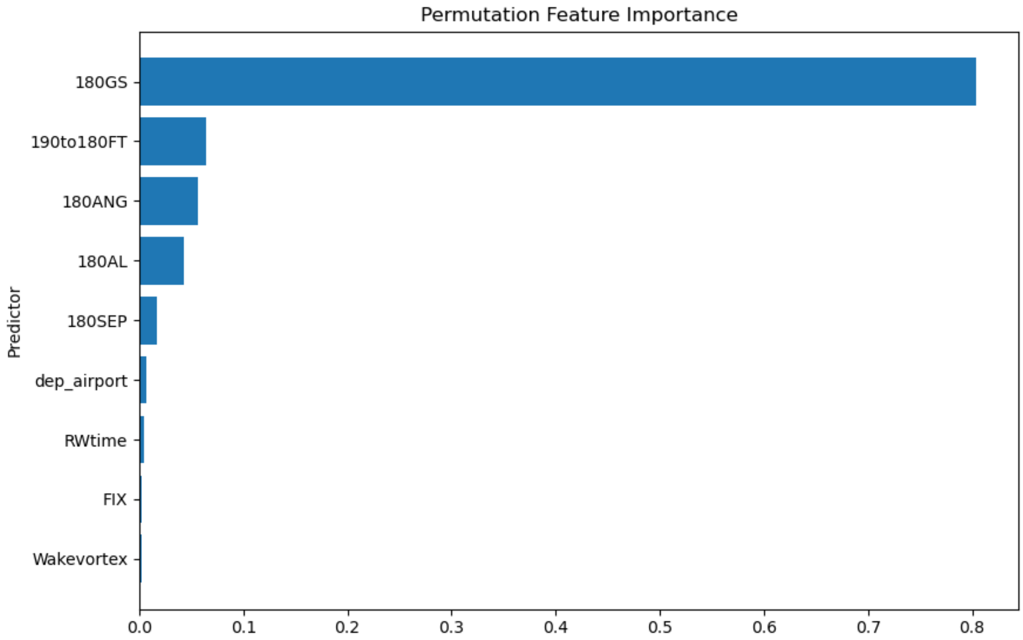

Figure 4 – Permutation Feature Importance Plot when predicting the flight time taken from 180 NM to 30 NM at 180 NM in a Random Forest by the Ohshima-centered model

Figure 9 represents the SHAP values when predicting the flight time from 180 to 30 NM at 180 NM in a Random Forest by the Ohshima-centered model.
Figure 9 shows that a high ground speed or high separation at 180 NM tends to make the flight time taken from 180NM to 30NM shorter, and the low altitude at 180NM tends to shorten the flight time. The categorical variables have smaller SHAP values than the numerical variables, but among them, the departure airport can have large values. According to it, flights having departed from Osaka(dep_airport=0) tend to have longer flight times.

6. Conclusion

We have presented several machine learning models such as elastic net and decision tree to predict flight times between the en-route airspace and the entrance to the terminal airspace. Our experiments verify that the GBDT achieved peak prediction accuracy, meaning the decision tree has a high affinity with air traffic management. We can improve the prediction error by not only changing machine learning methods, but also working out how to extract the feature. The Ohshima-centered model outperforms the Haneda-centered model by a max of 25% in RMSE. We also attempt to explain to stake holders like dispatchers, pilots, and air controllers why the model gives the prediction. Accordingly, we introduced four methods: PFI, PD, ICE and SHAP to further understand the prediction results. According to them, although numerical variables like ground speed at a certain point impact significantly on prediction, categorical variables such as time slot and departure airports also affects it. Some future work could involve selecting machine learning methods and appropriate airspace division to further improve prediction accuracy. We should also consider weather conditions like wind direction or speed because airplanes are highly affected by them.

7. Acknowledgements

This research was supported by a Grant-in-Aid for Scientific Research (20H04237). We would like to express our sincere appreciation for the data and technical support provided by the Civil Aviation Bureau of the Ministry of Land, Infrastructure, Transport and Tourism.

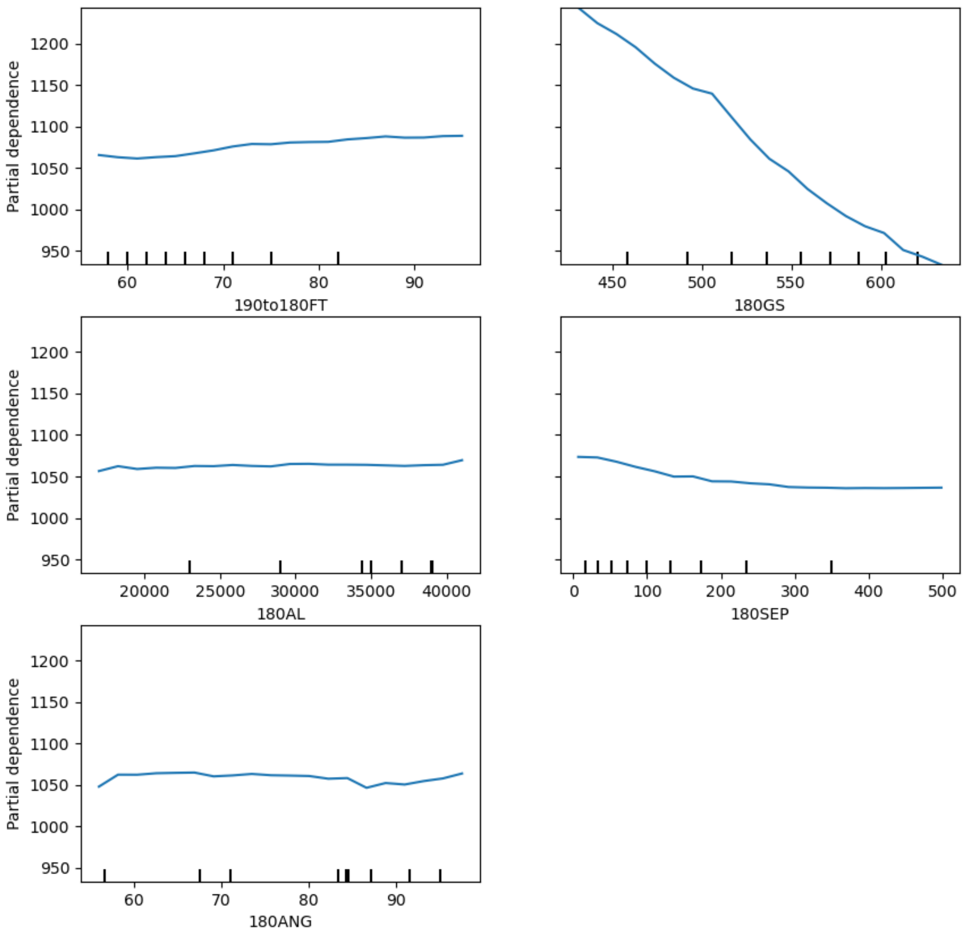

Figure 5 – PD Plot of the numerical variables

8. Contact Author Email Address

mailto: morikawa-nobuharu823@g.ecc.u-tokyo.ac.jp

9. Copyright Statement

References

[1] Japan Aircraft Development Corporation (JADC), [Market forecast for commercial aircraft 2020-2039] Minkan koukuuki ni kansuru shijou yosoku 2020-2039 (in Japanese), http://www.jadc.jp/files/topics/157_ext_01_0.pdf

[2] Itoh E and Miyazawa Y, A Study on the Stochastic Modeling of Traffic Flow Arriving in Tokyo International Airport, *the 56th Aircraft Symposium of Japan Society for Aeronautical and Space Science (JSASS)*, November 2018.

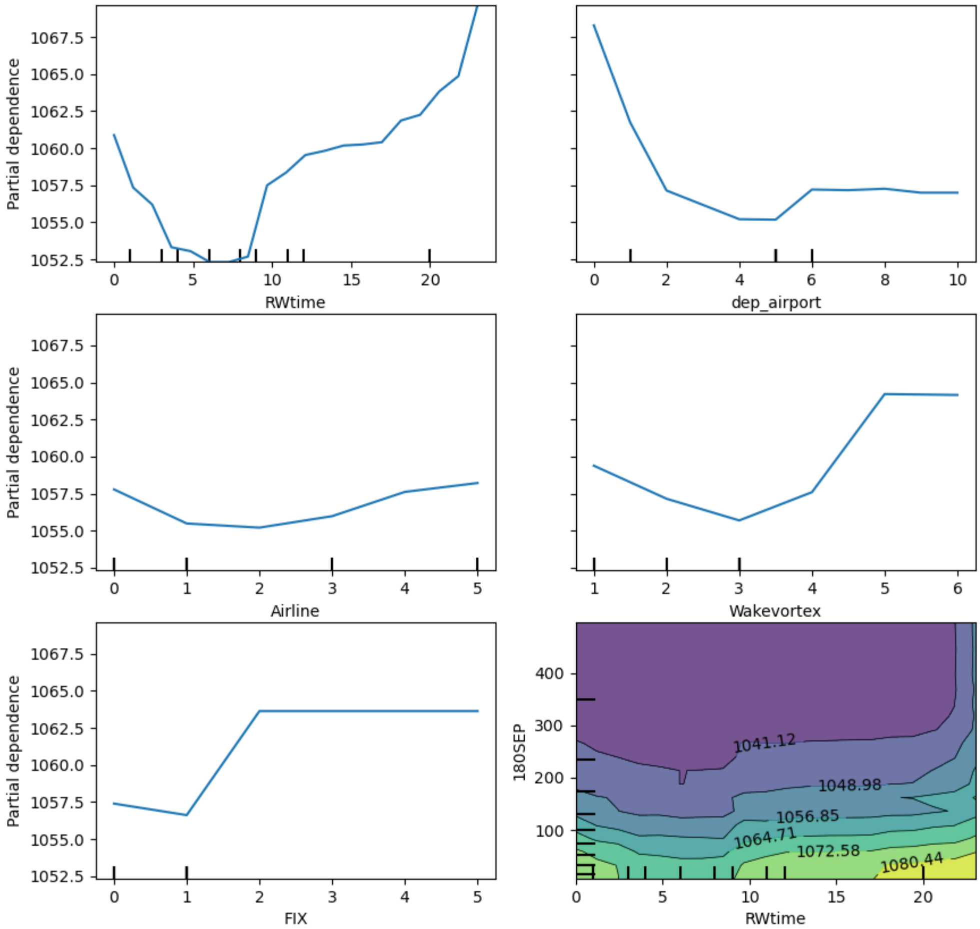

Figure 6 – PD Plot of the categorical variables

[3] Itoh E and Mitici M, Queue-Based Modeling of the Aircraft Arrival Process at a Single Airport. Aerospace. 2019, 6.10.103.

[4] Itoh E and Mitici M, Evaluating the Impact of New Aircraft Separation Minima and Available Airspace Capacity and Arrival Delay, *The Aeronautical Journal*, 124(1274), pp. 447-471, April 2020.

[5] Itoh E and Mitici M, Analyzing Tactical Control Strategies for Aircraft Arrivals Using Queue-based Modeling, *Journal of Air Transport Management*, 89(101938), October 2020.

[6] Itoh E, Miyazawa Y, Finke M, and Rataj J, Macroscopic Analysis to Identify Stage Boundaries in Multi-Stage Arrival Management, Air Traffic Management and Systems IV, *Springer nature*, 2020.

[7] Sekine K, Kato F, Kageyama K, and Itoh E, Data-driven Simulation for Evaluating the Impact of Lower Arrival Aircraft Separation on Available Runway and Airspace Capacity at Tokyo International Airport. *Aerospace*. 2021, 8(6), 165.

[8] Delahaye D, Puechmorel S, Alam S, and Féron E, 2017. Trajectory Mathematical Distance Applied to Airspace Major Flows Extraction.*Proceedings EIWAC 2017*.

[9] Olive X and Basora L. Identifying Anomalies in Past En-Route Trajectories with Clustering and Anomaly Detection Methods. *Proceedings of the 13th USA/Europe Air Traffic Management Research and Development Seminar*, Vienna, 2019.

[10] Olive X and Morio J. Trajectory Clustering of Air Traffic Flows around Airports. *Aerospace Science and Technology 84*, pp. 776–81, January 2019.

[11] Rebollo J and Balakrishnan H, Characterization and Prediction of Air Traffic Delays, *Transportation Re-*

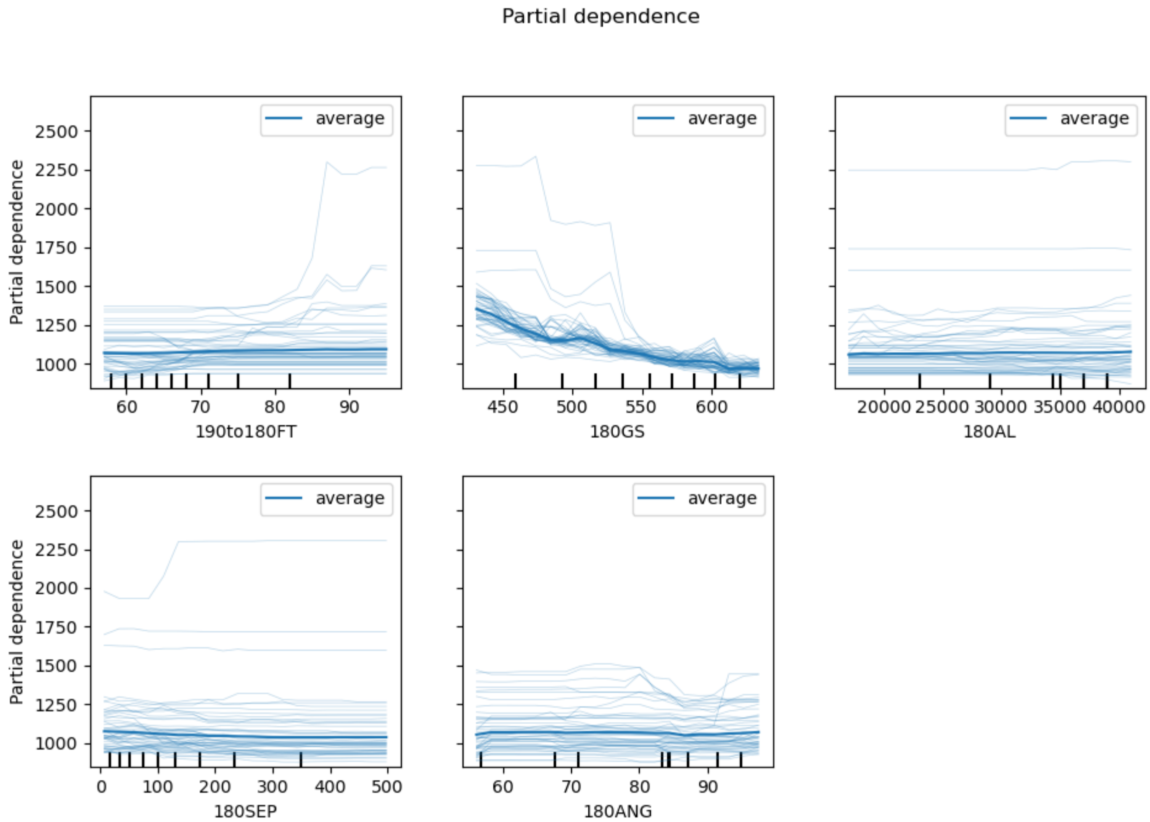

Figure 7 – ICE Plot of the numerical variables

search, Part C: Emerging Technologies*, pp. 231-241, 2014, 44:

[12] Ayhan S, Costas P, and Samet H, Predicting Estimated Time of Arrival for Commercial Flights, *KDD 2018*, London, United Kingdom, August 19-23, 2018.

[13] Gerdes I, Temme S, and Schultz M, Dynamic airspace sectorisation for flight-centric operations, *Transportation Research Part C: Emerging Technologies*, Vol. 95, 2018.

[14] Wang Z, Liang M, and Delahaye D, A hybrid machine learning model for short-term estimated time of arrival prediction in terminal maneuvering area, *Transportation Research Part C: Emerging Technologies*, 95: pp.280-294, 2018.

[15] Herrema F, Curran R, Hartjes S, Ellejmi M, Bancroft S and Schultz M, A machine learning model to predict runway exit at Vienna airport, *Transportation Research Part E: Logistics and Transportation Review*, 131. 2019.

[16] International Civil Aviation Organization (ICAO), The Procedures for Air Navigation Services – Air Traffic Management (PANS-ATM), Doc 4444, November 2020.

[17] Zou H and Hastie T, Regulation and Variable Selection via the Elastic Net, *Journal of the Royal Statistical Society*, Series B, 67(2), pp. 301-320, 2005.

[18] Segaran T, Programming Collective Intelligence, *O'Reilly Media (2007)*. ISBN 978-4-87311-364-7.

[19] Breiman L, Random Forests, *Machine Learning* 45, pp. 5-32 (2001).

[20] Bradley B, and Greenwell B, Gradient Boosting. *Hands-On Machine Learning with R. Chapman & Hall*, pp. 221–245 (2019). ISBN 978-1-138- 49568-5.

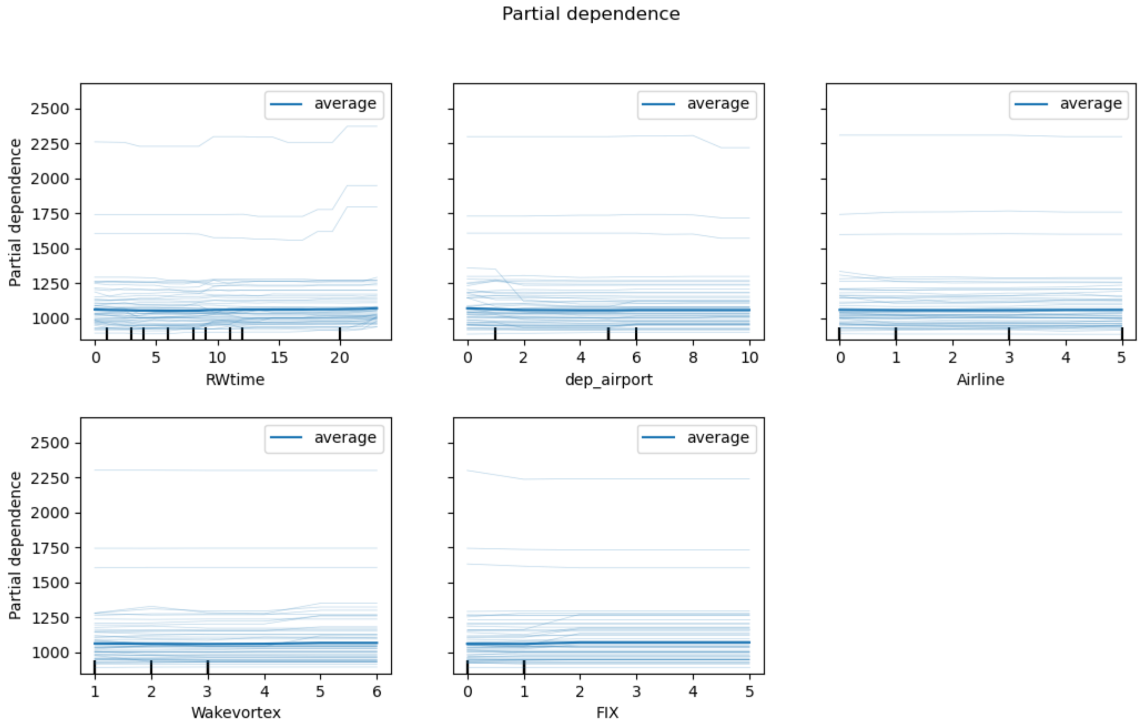

Figure 8 – ICE Plot of the categorical variables

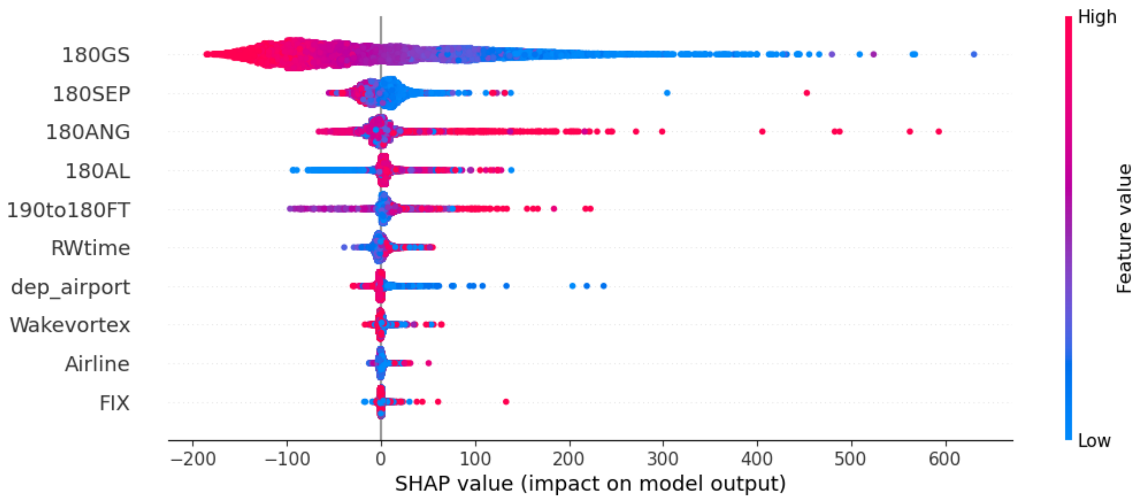

Figure 9 – SHAP Plot