ITM Web of Conferences **48**, 03005 (2022) https://doi.org/10.1051/itmconf/20224803005
ICCWCS 2022

Impact of Aircraft Performance and Time of the Day on Flight Arrival Delays Prediction in the United States: a Machine Learning Classification

Hajar Alla[1,2,3,*], *Lahcen* Moumoun[1,2,**], and *Youssef* Balouki[1,2,***]

[1]Laboratory of Mathematics, Computer and Engineering sciences. Mathematics and Computer Science Department
[2]Faculty of Science and Techniques Hassan First University of Settat. 26000 Settat, Morocco
[3]Corresponding Author

Abstract. The excessive growth of air traffic, with the limited airspace and airports capacity, results in a flight demand-capacity imbalance leading to air traffic delays. This paper explores the factors associated with delay in both microscopic and macroscopic ways. The aim is to develop a model which analyzes and predicts the occurrence of flight arrival delays using US domestic flight data for the year 2018. It will provide passengers, airlines and airport managers with reliable flight arrival schedules, and consequently reduce economic losses and enhance passengers trust. Beside database features, the proposed model is to the best of our knowledge the first attempt to predict flight arrival delays using three new features which are contributive factors to delays: Departure Time and Arrival Time of the day in which the flight was performed (Early morning, late morning, noon, afternoon, evening or night) and model of aircraft. Four Machine Learning classifiers namely Random Forest, Decision Trees, K-Nearest Neighbors and Naïve Bayes were used. In order to find the best parameters of each algorithm, we implemented Grid Search technique. The performance of each classifier was compared in terms of hyperparameters tuning, classification metrics and features description. The experimental results showed that the proposed system was able to predict flight arrival delays with the best Random Forest accuracy of 0.9356 and a higher number of correctly classified flights. To prove the importance of our findings, we compared our model to that of existing literature studies.

1 Introduction

Flight delays impact all aviation users such as travelers, airlines and airport authorities. It is considered as a dysfunction in the air traffic system. Bureau of Transportation Statistics (BTS) classifies flight delays in six categories: Air Carrier Delay, Aircraft Arriving Late, Security Delay, National Aviation System (NAS) Delay, Extreme Weather Delay, and Cancelled or Diverted flights. Figure 1 resumes the reasons of flight delays in all US major airports from January to December, 2017.

Among the factors influencing flight delays, we cite aircraft technical or mechanical problems, bad meteorological conditions, airport high density and capacity, late or disturbing passengers, insufficient airport infrastructure, delays in fueling, passengers embarking or disembarking, luggage loading or unloading and so on. Hence, traffic delays cost airlines million dollars per year comprising fuel costs due to fuel consumption, passenger delay costs paid to unhappy and complaining travelers and influence on next flights costs. To deal with this problem, researchers from all over the world conducted several studies to find solutions for traffic delays. In this study, a predictive classification model is proposed to estimate airline delays for United States study case using four effective Machine Learning algorithms such as Naïve Bayes, K-Nearest Neighbour, Decision Trees and Random Forest. On-time and delayed flights data for the year 2018 were extracted from BTS database.

Time of the day is an associated factor with delays. Indeed, depending on the season, airlines or travelers can pick the same part time of the day to take-in their flights (mornings, afternoons, evenings or nights). Choosing the same period can result in rush hours and an air and ground density.

According to [1], it is natural and inevitable to have differences in aircraft performances in air transportation environment. However, the differences do not only lead to congestion, density and delays but also cause economic losses and pollution of the environment through excessive gas emission and energy consumption. The delay increases in particular when slower aircraft models are followed by faster ones in a climb, cruise phase or even in the ground. For this reason and to the best of our knowledge, we proposed for the first time new features namely, Model of Aircraft, Departure Time and Arrival Time of the day. We will prove at the end of this study that the new proposed fea-

*e-mail: h.alla@uhp.ac.ma
**e-mail: lahcen.moumoun@uhp.ac.ma
***e-mail: youssef.balouki@uhp.ac.ma

tures were able to enhance the efficacy of our predictive model, accompanied with other features such as airport of origin and air carriers which are also contributive to flight delays.

Hyperparameters optimization is used in most studies to control the learning process, to get optimal and better values of the parameters and improve accuracy. In this paper, we chose Grid Search technique in order to find the best parameters for K-Nearest Neighbors, Random Forest and Decision Trees. We evaluated the performance of the proposed method based on hyperparameters tuning, classification metrics and features description. To prove the importance of our findings, we compared our model to that of existing literature studies.

The remainder of this paper is structured as follows: Section 2 describes existing literature on flight delay prediction. Section 3 presents in details the research methodology used in order to predict flight arrival delays. In Section 4, the performance of all classifiers is compared based on hyperparameters tuning, classification metrics and features description. Section 5 concludes the paper by summarizing the contributions of this study and recommending perspectives and further works.

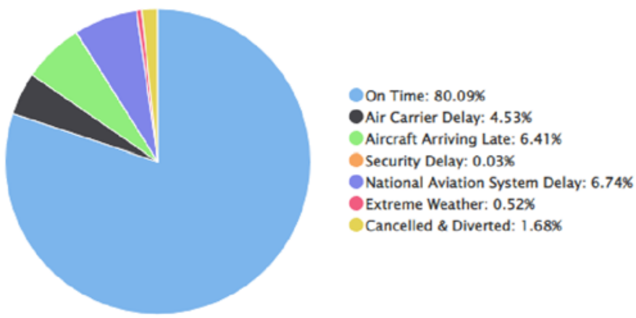

Fig. 1. Flight Delays by Cause - All Major Airports (January - December, 2017)

2 Literature Review

Due to its vital importance, the study of flight delays has taken more attention recently. Existing literature that studies flight arrival delay problems focuses the most on statistical and probabilistic techniques. Authors in [2] predicted flight delays by estimating taxi-out time using queuing model. [3] developed a probabilistic model based on delay distribution to estimate flights delays probability. [4] experimented several scenarios to estimate runway capacity of airports with Monte Carlo simulations. Authors in [5] focused on taxi-time reduction using Ant algorithm in order to optimize aircraft taxi movements. They proposed a model taking the shortest path of aircraft into account to reduce delays. Authors in [6] developed a system that allows a continuous taxiing to minimize taxi-time delays. Researchers in [7] attempted to create a statistical method to analyze departure and arrival flight delays. They developed a model for flight delays analysis using density functions. The authors claimed that Normal distribution performed better with departure delays while Poisson distribution worked better with en route and arrival delays.

Statistical methods are focused on probabilities and approximations which may lead to unsure conclusions and results. To cope with this issue, intelligent systems based on machine learning; which is a rapidly evolving field [8]; are hence needed. Authors in [9] focused on predicting flight arrival time at the block-off moment instead of airborne moment with linear regression and gradient boosting machines. Authors in [10] developed a model to forecast delays of US domestic flights based on multiple linear regression, decision trees and random forest algorithms. It has been concluded that Random forest outperformed the other models. To identify flight delay in advance, Authors in [11] described a predictive model using decision tree classifier which proved to be efficient. Decision Trees, Random Forest, AdaBoost and KNN were used as cost-sensitive classifiers to predict individual flight delays in [12]. [13] combined classification and regression. In the classification stage, Gradient Boosting Classifier performed the best and in the regression stage, Extra-Trees Regressor was the best performing classifier. Authors in [14] applied SVM, Gradient Boosting, Random Forest, AdaBoost and Decision Tree classifiers to predict on-time arrival flights. Random Forest was the best classifier with an accuracy of 77%. [15] established an improved Support Vector Machine (SVM) model to predict flight departure delay. The experimental results show that the proposed model is able to predict flight delay time with a high effectiveness and accuracy than other algorithms. In order to analyze and predict flight arrival delays, authors in [16] applied four supervised machine learning classifiers namely random forest, Support Vector Machine (SVM), Gradient Boosting Classifier (GBC) and k-nearest neighbour algorithm. The best classifier proved to be gradient boosting with an accuracy of 79.7%. Authors in [17] developed a model to estimate and predict flight arrival delay using decision tree, logistic regression and neural networks algorithms. The performance was approximately equal to 91% for all three classifiers.

In the existing studies, researchers have addressed flight delays from different perspectives. In this paper, we developed a predictive model that studies several factors influencing flight delays in the macroscopic level. Moreover, to the best of our knowledge, new microscopic features were proposed and considered in order to enhance the accuracy of the flight prediction system. The model is trained using various supervised machine learning classifiers that have demonstrated to be efficient in previous studies. The proposed system guarantees parameters tuning for a better performance.

3 Proposed Methodology

3.1 Research Motivation

Delay is the second most significant factor to consider in aviation environment, after safety. It annoys passengers and airport managers and increase airlines economic

losses. Since it is still an important ongoing research area, multiple studies have been conducted to cope with this issue. A system, which predicts flight arrival delays and provides airport personnel, companies, passengers and aviation users with the delay before its occurrence, is needed. Several macroscopic factors impact flight delays such as bad weather conditions, severe crosswinds [18], technical problems, late and disturbing passengers, airport crowdedness, runway queues, lack in airport infrastructure, aircraft maintenance delays, flight check-list delays, etc.
Our objective in this study is to identify and use new microscopic factors which participate in flight delays to develop a predictive model based on good performing machine learning classifiers. Figure 2 describes the process followed to create and generate the proposed method. Flights data are extracted from BTS database. ICAO[1] aircraft models, manufacturers and registrations are imported from Github. In order to remove redundant, duplicated, noisy and unnecessary information, preprocessing and cleaning techniques are applied. Only relevant features that are contributive to flight delays need to be selected. Hence, a features selection is necessary. In order to improve the efficacy of the proposed system, other microscopic factors of delays are considered. For that and to the best of our knowledge, three new features; namely Departure Period of the day, Arrival Period of the day and Model of aircraft; are created and inserted into the model. A data split of 70:30 ratio is applied generating training and testing data. To generate the best parameters of all four classifiers, Grid Search technique is utilized for parameters tuning. The performance of the proposed predictive model is finally evaluated using classification metrics.

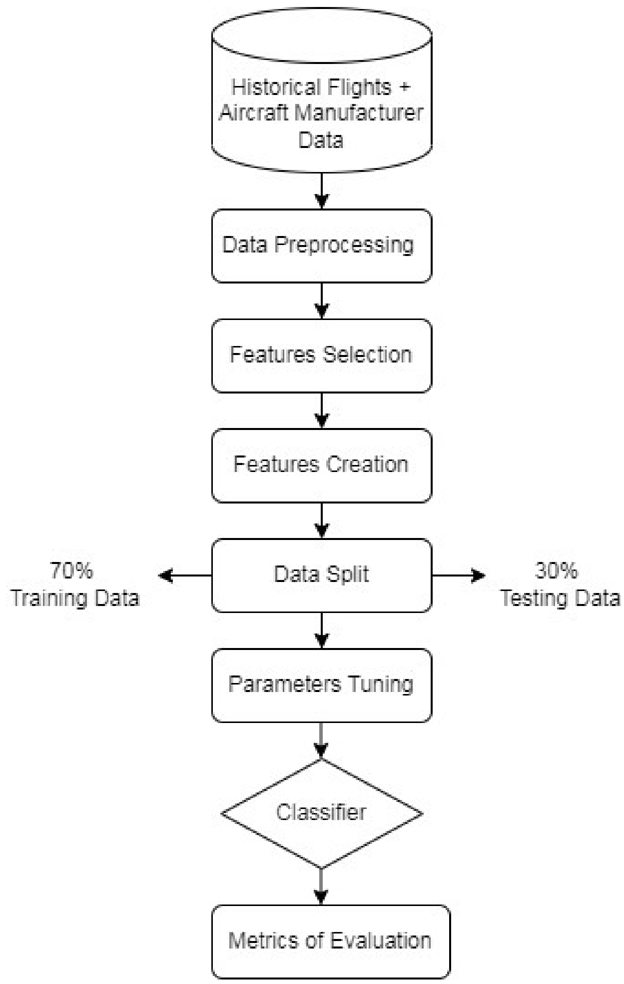

Fig. 2. Flowchart of the proposed method

3.2 Data Collection

United States on-time performance data from 1st of January to 31st of December 2018 are achieved from the Bureau of Transportation Statistics database (BTS) located in the Department of Transportation in The United States. The dataset contains about 94810 recordings of domestic flights. The accessibility of data saved since 1987, reliability, transparency and independence from political influence [19] encouraged us to choose BTS as the source of our data collection. According to BTS, all flights whose difference between scheduled and actual arrival times is 15 min or higher are considered delayed. Aircraft manufacturer, models and registration were extracted from Github[2], saved as aircraft performance and merged with BTS flight movements to create a unique dataset. The file extracted contained about 353043 instances of aircraft models and manufacturers.

3.3 Data Preprocessing

Real-world data are often incomplete, incorrect and mistaken. It may be also noisy and inconsistent [20]. Since machine learning algorithms work better with numerical variables, data cleaning need to be accomplished. Duplicate data, missing and null values need to be erased. Categorical data are converted to numeric.

3.4 Features Analysis

The data extracted contained several attributes and data fields. Only features which have an impact on flight delays were selected and unnecessary attributes were removed.

Carrier and Company Delay BTS defines a carrier or company delay as a delay or cancellation caused due to circumstances within the airline's control (e.g. maintenance or crew problems, slower aircraft cleaning, baggage loading, fueling, etc.). Due to repeated delays, passengers may change an airline for another that is less subject to delays. Furthermore, a failure in turnaround process caused by push-back, fueling/refueling, catering or mechanical check operations may engender delays. Personnel stress and fatigue [19] or personnel labor strike [21] are considered contributive to carrier traffic delays.

Origin and Destination Airports The distance between the origin and destination airports determine the duration of the flight. If it is higher, the flight is likely to be late. [21] concluded that the longer is the distance, the higher is the delay.

[1] ICAO: International Civil Aviation Organization
[2] https://github.com/RobAltenburg

Departure Delay Departure delays are contributive to arrival delays. [22] claimed that congestion at destination airport is mostly originated at the departure airport. A traffic experiencing a delay on departure may be also delayed on arrival [21]. Hence, departure delay is a relevant feature to our study.
To improve the efficiency of our system, we created three other features,namely Aircraft Model, Departure Period and Arrival Period of the day in which the flight was performed.

Aircraft Model Some aircraft can be more flexible in arriving on-time depending on the performance [23]. According to [24], aircraft performance differences cause flight delays. Lower performance aircraft determine airspace capacity. If a low performance aircraft precedes a higher performance one, it may cause a delay by retarding the following flight since the speed of the constraining aircraft is lower. The high performance aircraft is then forced to maintain the same speed or lower which causes traffic delays. In Table 1, we present, from our dataset, the remarkable difference in delays percentage depending on aircraft model and manufacturer. As we can notice, 38% from the flights performed by BOEING 7378FH were delayed while only 11% were not on-time in the case of MCDONNELL DOUGLAS AIRCRAFT MD-88. Hence, the difference in performance has an impact on traffic delays and the model of aircraft is a necessary feature in our study.

Period of the Day As we know, the period of a day is an important factor to consider when traveling. In fact, every passenger has a preferable period to take in a flight. Some prefer mornings, others choose to travel by night, etc. Also, there is some peak or rush period of the day in which traffic tends to be dense leading to flight delays. In general, we account more flights in mornings and afternoons than nights due to the visibility. For this purpose, we decided, in our work, to create two new features, namely Departure Period and Arrival Period of the day which are respectively, the time of the day in which a flight has departed and the time of the day in which a flight has arrived to its destination. According to the world's number one website on time database[3], we decided to take 06:05 US Local Time as the minimum sunrise time and 21:06 US Local Time as the maximum sunset time to cover up all the year 2018. Table 2 displays the period of the day depending on the interval of hours. Each interval is affected a label.

3.5 Supervised Learning Classifiers

In order to predict whether a flight will be delayed or not, four supervised classifiers were applied namely KNearest-Neighbors, Naïve Bayes, Decision Trees and Random Forest.
K-Nearest-Neighbors is a classifier which take into consideration k adjacent points as neighbors in order to classify a point x. The process followed by k-NN classifier is described as follows [25]:

- Calculate the number of adjacent neighbors k
- Determine the distance from the point x to all the neighboring samples
- Sort the distance and find the nearest neighbors by measuring the shortest Euclidean distance
- Collect the label of each nearest neighbor. The label with the majority vote are the one assigned for the prediction

Naïve Bayes is a simple Bayesian network with multiple children (observed nodes) and a single parent (unobserved node) based on an assumption that all the variables are independent of each other. In a study carried out by [26], Naïve Bayes has proven to be more tolerant to noise, missing values, overfitting, irrelevant, redundant and interdependent attributes such as parity problems. It has a high speed of learning with respect to number of instances and attributes. Naïve Bayes is faster than k-NN at classification.
In Decision Trees algorithm, instances are classified and sorted based on the value of each feature which represents a node. The branches contain all possible values of the chosen node. All data are stored in the root node which is the very top of the tree. Each node asks a yes or no question and as an answer the data is subdivided into two subsets. Internal nodes have arrows pointing to them and they have arrows pointing away from them. Leaf nodes have arrows pointing to them but there are no arrows pointing away from them. Each internal node designates a test on an attribute, each branch designates an outcome of the test, and each leaf node (terminal node) holds a class label. Decision tree has the advantage of adaptation with data collected, a high speed of classification, a good accuracy and being able to deal with discrete, binary and continuous attributes [26]. Authors in [27] used decision tree algorithm on crime data to predict the probability of low, medium or high violent crimes in a country.
Random Forest is an algorithm capable of performing both regression and classification tasks. It consists of an ensemble of several individual decision trees. The more trees are in the forest, the more robust is the prediction. The trees are created randomly from input features of the training dataset. The required prediction is then generated by using Bagging technique. [28] considered Random Forest as one of the most successful ensemble learning techniques to be very popular and powerful. [26] rated Random Forest among classification algorithms with a high percentage of correctly classified parameters.

3.6 Hyperparameters Tuning

To find optimal hyperparameters for Random Forest, Decision Trees and K-Nearest Neighbors and improve accuracy, Grid Search technique was chosen over Manual Search with 10-fold cross validation. For the choice of hyperparameters, we were inspired from a study conducted by [29]. We evaluate all the possible combinations of parameters and keep the best combination. The best parameters of Random Forest were: 'max_features:

[3]United States, USA – Sunrise, Sunset, and Daylength, 2018, https://www.timeanddate.com/

Table 1. Delay percentage of the most common aircraft types in the dataset

Model	Manufacturer	Percentage of delay
737-8FH	BOEING	38%
777-323ER	BOEING	33%
H101 "SALTO"	START AND FLUG GMBH	30%
DOMINATOR	BENHAM JOHN	28%
R66	ROBINSON HELICOPTER CO	28%
340A	CESSNA	27%
SR22	CIRRUS	27%
DA 42	DIAMOND AIRCRAFT IND GMBH	26%
FALCON 900 EX	DASSAULT	26%
A321-231	AIRBUS	25.5%
GVI (G650)	GULFSTREAM AEROSPACE CORP	25%
A320-214	AIRBUS	24%
ERJ 190-100 IGW	EMBRAER	24%
AT-602	AIR TRACTOR INC	22%
182R	CESSNA	19%
B739	BOEING	13%
MD-88	MCDONNELL DOUGLAS AIRCRAFT CO	11%

Table 2. US Part-Time of the Day by Local Time in 2018

Interval of hours	Time of the Day Label
From 00:00 to 06:05	Early Morning
From 06:06 to 11:59	Late Morning
At 12:00	Noon
From 12:01 to 17:59	Afternoon
From 18:00 to 21:06	Evening
From 21:07 to 23:59	Night

sqrt' and 'n_estimators: 700' with the highest accuracy of 0.936344. In case of Decision Trees, 'criterion': 'entropy', 'max_depth': 20 and 'max_leaf_nodes': 150 were the best parameters with an accuracy of 0.923033. In K-Nearest Neighbors algorithm, 'metric': 'minkowski', 'n_neighbors': 15 and 'p': 1 were the best parameters with an accuracy of 0.811090.

4 Results and Discussions

The experiments were conducted on a dataset of 94810 flights and 353043 models of aircraft. After the files concatenation and data cleaning, a final dataset of 58971 recordings was utilized. Training and testing data were generated using a 70:30 ratio split. In this study, three new features namely Departure Period of the day, Arrival Period of the day and Model of aircraft were created in order to analyze new factors contributing in flight delays and achieve a good prediction. To create the model, K-Nearest Neighbors, Naives Bayes, Random Forest and Decision Trees were applied. To maximize the accuracy and obtain a better classification, parameters were tuned using Grid Search technique.

4.1 Features Discussion

4.1.1 Period of the Day

We extracted some of the features from the database but created other relevant ones: Departure period of the day and Arrival Period of the day duly explained in the methodology section. By analyzing our data, we found that the most delayed flights on departure were performed on late mornings, afternoons and evenings. Morning is perfect for travelling since it is the least crowded time in the day with generally a calm weather that passengers prefer to take a flight in. Afternoons and evenings are subject to rush hour, bad weather conditions and traffic density which cause delays. At arrival, the flights are also delayed at night maybe because the delay propagation in afternoons and evenings affect also the night period. Finally, we notice that the less affected by delays were the flights performed at noon which are quasily non-existent according to Figure 3 and Figure 4.

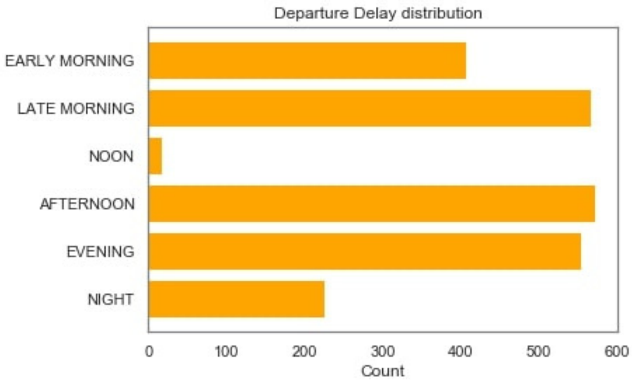

Fig. 3. Departure delay by part of the day

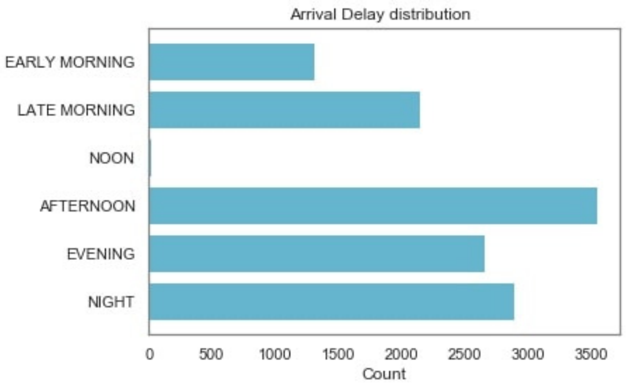

Fig. 4. Arrival delay by part of the day

4.1.2 Model of Aircraft

In Figure 5, we plot the percentage of the 5-most delayed aircraft model in our database. We notice that Airbus A321-231, A320-232 and EMBRAER ERJ 190-100 IGW were the most affected by arrival delays. The possible reason of the delay may be not being able to handle bad weather conditions.

4.1.3 Carrier Delay

We have considered also the company delay among our features. According to Figure 6, JetBlue (B6) is the most delayed company at arrivals while Alaska Airlines (AS) is the most arriving on-time company. Airlines delay may be caused by carrier issues such a slow push-back operations, mechanical check lists, fuel service, loading luggage, personnel labour strike, etc.

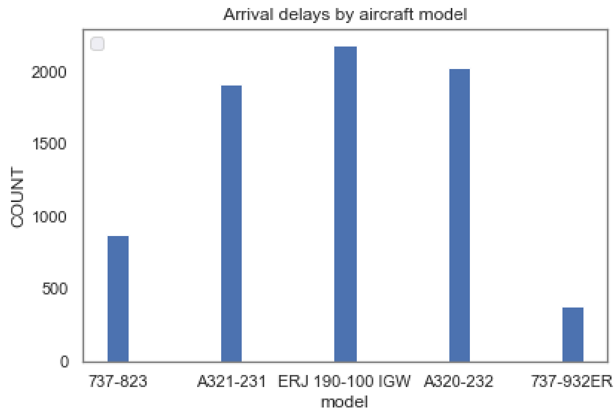

Fig. 5. Arrival delay by model of aircraft

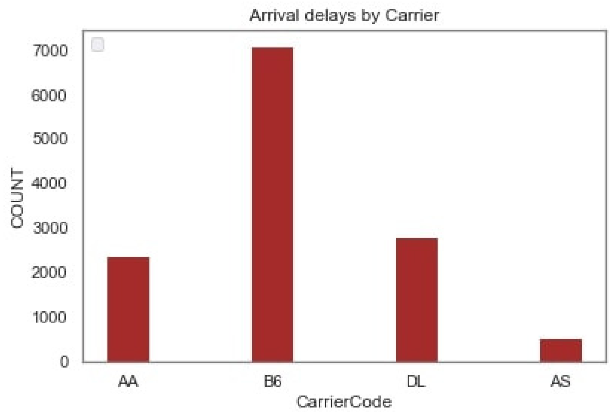

Fig. 6. Arrival delay by Carrier

4.1.4 Airport of Origin

From 63 origin airports used in this study, we plot only the 10-most delayed ones in Figure 6. We notice that Los Angeles International Airport (LAX), San Francisco International Airport (SFO) and Orlando International Airport (MCO) are respectively the 3-most delayed airports in US 2018 based on our features analysis.

4.2 Impact of Hyperparameters Tuning

In order to improve the accuracy and find optimal parameters, this study conducted tuning parameters of Random Forest, Decision Trees and K-Nearest Neighbors with Grid Search technique. We compare in Table 3 and Table 4 the performance of the algorithms with and without tuning, respectively.
We notice that the evaluation metrics of the three algorithms were improved when the tuning was applied. The experimental results show that parameters tuning has successfully generated the best classifier which is Random Forest with an accuracy of 0.9356 and a higher number of correctly classified flights.

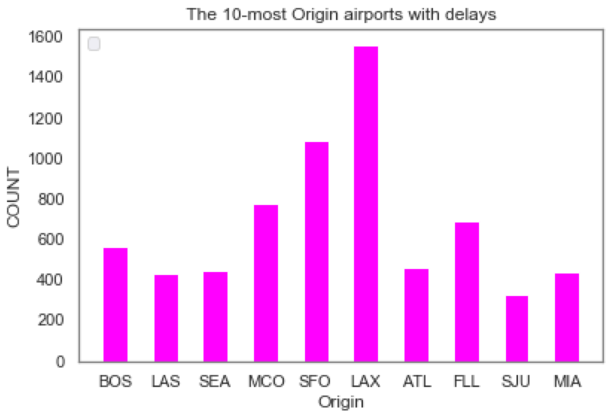

Fig. 7. The10-most delayed airports of departure

Table 3. Evaluation metrics with tuning

Algorithm	Accuracy	Precision		Recall		F1-score	
		0	1	0	1	0	1
Knn	0.81	0.75	0.90	0.92	0.69	0.83	0.78
DT	0.89	0.88	0.91	0.91	0.88	0.90	0.89
RF	0.93	0.92	0.95	0.95	0.92	0.93	0.93

Table 4. Evaluation metrics without tuning

Algorithm	Accuracy	Precision		Recall		F1-score	
		0	1	0	1	0	1
Knn	0.78	0.77	0.80	0.81	0.76	0.79	0.78
DT	0.87	0.86	0.87	0.87	0.86	0.87	0.87
RF	0.90	0.88	0.93	0.93	0.88	0.90	0.90

4.3 Models Benchmark

To prove that the experimental results of our model are fair and significant, we compare, in Table 5, the accuracy of our proposed Random Forest model to that of other existing literature results. To predict arrival delays, Jun Chen [30] chose an optimal feature selection in a Random Forest model than directly using all the features of available datasets. Jiage Huo et al. [31] developed a Random Forest system to predict flight delays for airlines in Hong Kong based on features selection. Chakrabarty et al. [16] utilized Random Forest data mining to analyze and estimate arrival delay of the flights. Gui et al. [32] implemented a random forest-based model to predict individual flight delay.

We notice that, using new features namely Departure Time of the Day, Arrival Time of the Day and Model of Aircraft accompanied with other features such as Airport of Origin, and Air Carrier Delays, generated a high accuracy model compared with existing studies.

Table 5. Accuracy benchmark of our proposed model and that of existing studies

	The proposed model	[30]	[31]	[16]	[32]
Accuracy	0.9356	0.8672	0.6639	0.7870	0.9020

5 Conclusion

Flight delay has been an important and challenging research subject as air demand grows year after year. In response to growth concerns, researchers analyzed flight delays from different perspectives. This study focused on flight arrival delays prediction using four Machine Learning classification algorithms namely Random Forest, Decision Trees, KNN and Naïve Bayes. In order to enhance the performance of the proposed prediction model, we created and added three new features: Departure period of the day, Arrival period of the day and model of aircraft. Parameters tuning using Grid Search Technique was performed in order to obtain an improved system. The model has achieved the highest accuracy of 0.9356 with Random Forest.

The proposed model can be used by airports managers, airlines and travelers to predict flight arrival delays of United States airports. It can be utilized also as a decision support tool for air traffic controllers. In fact, if they are informed of flight arrival times and possible delays in advance, they can be aware of peak hours of flights and whether it is necessary or not to add more controllers to the shift team. As a limitation, the dataset contains only on-time data for non-stop domestic flights. In future works, we consider extending the dataset to include also stop-over and international flights. Furthermore, it will be interesting to add more features and use deep learning in order to enhance the accuracy of the predictive model.

References

[1] Cavcar, Aydan, and Mustafa Cavcar. "Impact of aircraft performance differences on fuel consumption of aircraft in air traffic management environment." Aircraft Engineering and Aerospace Technology (2004)

[2] Idris, Husni, John-Paul Clarke, Rani Bhuva, and Laura Kang. Queuing model for taxi-out time estimation. (2001).

[3] Tu, Yufeng, Michael O. Ball, and Wolfgang S. Jank. "Estimating flight departure delay distributions—a statistical approach with long-term trend and short-term pattern." Journal of the American Statistical Association 103, no. 481, 112-125, (2008).

[4] Zhong, Z. W., Dhir Varun, and Y. J. Lin. "Studies for air traffic management R&D in the ASEAN-region context." Journal of Air Transport Management 64, 15-20, (2017).

[5] Nogueira, Kamila B., Paulo HC Aguiar, and Li Weigang. "Using ant algorithm to arrange taxiway sequencing in airport." International Journal of Com-

puter Theory & Engineering 6, no. 4, 857-361, (2014).

[6] Jiang, Yu, Xinxing Xu, Honghai Zhang, and Yuxiao Luo. "Taxiing route scheduling between taxiway and runway in hub airport." Mathematical Problems in Engineering 2015 (2015).

[7] Mueller, Eric, and Gano Chatterji. "Analysis of aircraft arrival and departure delay characteristics." In AIAA's Aircraft Technology, Integration, and Operations (ATIO) 2002 Technical Forum, p. 5866. (2002).

[8] Oualid, Adil, Abderrahim Hansali, Youssef Balouki, and Lahcen Moumoun. "Application of Machine Learning Techniques for Credit Risk Management: A Survey." In The International Conference on Information, Communication & Cybersecurity, pp. 180-191. Springer, Cham, (2021).

[9] Achenbach, Anna, and Stefan Spinler. "Prescriptive analytics in airline operations: Arrival time prediction and cost index optimization for short-haul flights." Operations Research Perspectives 5, 265-279, (2018).

[10] Kalliguddi, Anish M., and Aera K. Leboulluec. "Predictive modeling of aircraft flight delay." Universal Journal of Management 5, no. 10, 485-491, (2017).

[11] Katpadi, Vellore, and Tamil Nadu. "FLIGHT DELAY PREDICTION USING SUPERVISED MACHINE LEARNING."

[12] Choi, Sun, Young Jin Kim, Simon Briceno, and Dimitri Mavris. "Cost-sensitive prediction of airline delays using machine learning." In 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC), pp. 1-8. IEEE, (2017).

[13] Thiagarajan, Balasubramanian, Lakshminarasimhan Srinivasan, Aditya Vikram Sharma, Dinesh Sreekanthan, and Vineeth Vijayaraghavan. "A machine learning approach for prediction of on-time performance of flights." In 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC), pp. 1-6. IEEE, (2017).

[14] Etani, Noriko. "Development of a predictive model for on-time arrival flight of airliner by discovering correlation between flight and weather data." Journal of Big Data 6, no. 1, 1-17, (2019).

[15] Wu, Weinan, Kaiquan Cai, Yongjie Yan, and Yue Li. "An improved svm model for flight delay prediction." In 2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC), pp. 1-6. IEEE, (2019).

[16] Chakrabarty, Navoneel, Tuhin Kundu, Sudipta Dandapat, Apurba Sarkar, and Dipak Kumar Kole. "Flight arrival delay prediction using gradient boosting classifier." In Emerging technologies in data mining and information security, pp. 651-659. Springer, Singapore, (2019).

[17] Kuhn, Nathalie, and Navaneeth Jamadagni. "Application of machine learning algorithms to predict flight arrival delays." CS229 (2017).

[18] Jarrar, Abdessamad, Abderrahim Ait Wakrime, and Youssef Balouki. "Formal approach to model complex adaptive computing systems." Complex Adaptive Systems Modeling 8, no. 1 (2020): 1-35.

[19] Alla, Hajar, Lahcen Moumoun, and Youssef Balouki. "Flight arrival delay prediction using supervised machine learning algorithms." In Intelligent Systems in Big Data, Semantic Web and Machine Learning, pp. 231-246. Springer, Cham, (2021).

[20] Moreira, Leonardo, Christofer Dantas, Leonardo Oliveira, Jorge Soares, and Eduardo Ogasawara. "On evaluating data preprocessing methods for machine learning models for flight delays." In 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1-8. IEEE, (2018).

[21] Alla, Hajar, Lahcen Moumoun, and Youssef Balouki. "A multilayer perceptron neural network with selective-data training for flight arrival delay prediction." Scientific Programming 2021 (2021).

[22] Oza, Sruti, Somya Sharma, Hetal Sangoi, Rutuja Raut, and V. C. Kotak. "Flight delay prediction system using weighted multiple linear regression." International journal of engineering and computer science 4, no. 4, 11668-11677, (2015).

[23] Tielrooij, Maarten, Clark Borst, Marinus M. Van Paassen, and Max Mulder. "Predicting arrival time uncertainty from actual flight information." In Proceedings of the 11th USA/Europe Air Traffic Management Research and Development Seminar, pp. 577-586. (2015).

[24] Cavcar, Aydan, and Mustafa Cavcar. "Impact of Aircraft Performance Characteristics on Air Traffic Delays." Turkish Journal of Engineering and Environmental Sciences 28, no. 1, 23-24, (2004).

[25] Priyanka, G. "Prediction of airline delays using K-nearest neighbor algorithm." Int. J. Emerg. Technol. Innov. Eng 4, no. 5, 87-90, (2018).

[26] Osisanwo, F. Y., J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi, and J. Akinjobi. "Supervised machine learning algorithms: classification and comparison." International Journal of Computer Trends and Technology (IJCTT) 48, no. 3, 128-138, (2017).

[27] Ivan, Niyonzima, Emmanuel Ahishakiye, Elisha Opiyo Omulo, and Danison Taremwa. "Crime Prediction Using Decision Tree (J48) Classification Algorithm." (2017).

[28] Azar, Ahmad Taher, Hanaa Ismail Elshazly, Aboul Ella Hassanien, and Abeer Mohamed Elkorany. "A random forest classifier for lymph diseases." Computer methods and programs in biomedicine 113, no. 2, 465-473, (2014).

[29] Ramadhan, Muhammad Murtadha, Imas Sukaesih Sitanggang, Fahrendi Rizky Nasution, and Abdullah Ghifari. "Parameter tuning in random forest based on grid search method for gender classification based on voice frequency." DEStech Transactions on Computer Science and Engineering 10 (2017).

[30] Chen, Jun, and Meng Li. "Chained predictions of flight delay using machine learning." In AIAA Scitech 2019 forum, p. 1661. 2019.

[31] Huo, Jiage, Kin Lok Keung, Carman Ka Man Lee, Kam KH Ng, and K. C. Li. "The Prediction of Flight Delay:Big Data-driven Machine Learning Approach." In 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), pp. 190-194. IEEE, 2020.

[32] Gui, Guan, Fan Liu, Jinlong Sun, Jie Yang, Ziqi Zhou, and Dongxu Zhao. "Flight delay prediction based on aviation big data and machine learning."IEEE Transactions on Vehicular Technology 69, no. 1, pp. 140-150, (2019).