Towards Flight Delays Reduction: The Effect of Aircraft Type and Part of Day on Arrival Delays Prediction

HAJAR ALLA[1] , LAHCEN MOUMOUN[2], YOUSSEF BALOUKI[3]

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

Summary
The basic objective of this study is to develop a model which analyzes and predicts the occurrence of flight arrival delays in the United States. Macroscopic and microscopic delay factors are discussed. In this research, we proposed new features which, to the best of our knowledge, were never used in previous studies, namely departure Period and Arrival Period of the day (Mornings, Afternoons, Evenings, Nights) and type of aircraft. US domestic flight data for the year 2018,extracted from Bureau of Transportation Statistics (BTS), were adopted in order to train the predictive model. We used efficient Machine Learning classifiers such as Naive Bayes, Decision Trees, K-Nearest Neighbors and Random Forest. To overcome the issue of imbalanced data, sampling techniques were performed. We chose Grid Search technique for best parameters selection . We evaluated the effectiveness of each algorithm by comparing performance metrics, parameters optimization, data balancing and features selection. As a result, Random Forest proved to be the best classifier with an accuracy of 93.56% and a well satisfying classification. The performance of each classifier was compared in terms of evalua- tion metrics, parameters tuning, data sampling and features selection. The experimental results showed that tuning and sampling techniques have successfully generated the best classifier which is MLP with an accuracy of 98.42% and a higher number of correctly classified flights.

Keywords:
Machine Learning Classification, Flight Delay Prediction, Multilayer Perceptron, Random Forest, Decision Trees, K-Nearest Neighbors

1. INTRODUCTION

This guide provides details to assist authors in preparing a paper for publication in JATIT so that there is a consistency among papers. These instructions give guidance on layout, style, illustrations and references and serve as a model for authors to emulate. Please follow these specifications closely as papers which do not meet the standards laid down, will not be published. Flight delays affect passengers, airlines and airport managers. Indeed, it is a major preoccupation in air transportation systems. According to Bureau of Transportation Statistics (BTS), in 2017, 19.07\% of all domestic flights in United States were delayed. Many factors are responsible for traffic delays including bad weather conditions, technical problems, late passengers, airport crowdedness, runway queues, lack in airport infrastructure, aircraft turnover delay (fueling/refueling, loading/unloading, boarding/disembarking, etc.) and late arriving aircraft which impacts the following flight by creating a delay propagation. All these factors result in airlines economic losses and penalties, flight cancellations, passengers complaints and dissatisfaction, delay propagation on next flights, fuel consumption and gas emissions. To solve flight delay issues, studies and researches were conducted to find solutions for traffic delays. The objective of this study is to predict flight arrival delays using efficient Machine Learning classification algorithms such as Multilayer Perceptron, Random Forest, Decision Trees and K-Nearest Neighbors(K-NN). Traffic information of US domestic airlines flights for the year 2018 were extracted from BTS database. The period of the day is a contributing factor to delays. In fact, depending on the season, airlines or passengers may choose the same period time of the day to perform their flights, it is either on night, evening , afternoon or morning, which may lead to a peak period causing density and traffic delays. Models of aircraft do not perform the same. Ones are sensitive to bad weather conditions, others have a low rate of climb or descent or a low velocity. Aircraft with more seats are more subject to flight delays since it transports more passengers that can be late for boarding which causes delays. For this purpose and to enhance the performance of the proposed prediction model, we created new other features that, to the best of our knowledge, were not used in previous studies, namely, Aircraft Model, Departure Period and Arrival Period of the day. through our work, we will demonstrate that airports of origin and air carriers are also responsible for flight delays. The problem of imbalanced data in classification has become more challenging recently. To handle this issue, under-sampling and over-sampling techniques are often employed. Synthetic Minority Oversampling Technique (SMOTE) and Tomek Links are used for respectively over-

Manuscript received November 5, 2024
Manuscript revised November 20, 2024
https://doi.org/10.22937/IJCSNS.2024.24.11.21

sampling and under-sampling imbalanced data. To improve the efficacy of our method, we utilized a combination of both techniques and applied a SMOTE-Tomek for balancing our dataset. Parameters tuning is implemented in machine learning to get optimal values of the parameters and reach a higher level of accuracy. In this research, we used Grid Search technique to find the best parameters for Random Forest, Decision Trees and K-NN. We examine and compare the performance of the proposed method based on classification metrics, parameters tuning, data sampling and features selection.

The rest of the paper is organized as follows: Section 2 provides a review on existing literature of flight delay prediction. Section 3 describes in details the methodology proposed to develop the predictive model. In Section 4, we compare the performance of all classifiers in terms of evaluation metrics, parameters tuning, data balancing and features selection. Section 5 concludes the research and recommends perspectives and further works.

2. LITERATURE REVIEW

Nowadays, more interest has been given to flight delays because of the massive increase in air traffic. Existing studies in this area concentrate the most on statistical methods based on approximations and probabilities leading to unsure conclusions and results. [1] focused on taxi-out delays prediction for aircraft based on a queuing model. [2] described and studied major factors that influence flight departure delays by estimating the entire delay distribution. [3] performed an investigation about flight delays based on Monte Carlo simulations considering runway occupancy time, separation between aircraft and other operational scenarios at the airport. Authors in [4] proposed solutions in order to optimise taxi-scheduling operations for aircraft with Ant Colony Optimization. the model shows a performance of 32% in taxiing time minimisation, aircraft collision conflicts included. The objective of the study conducted by authors in [5] is to reduce both the total aircraft taxiing and waiting times using genetic algorithm. Authors in |6] studied departure, en route and arrival flight delays using historical data of United States airports based on density functions modeling.

Unlike the traditional statistical methods which generally have proven to be weak, slow, and limited, machine learning algorithms become more popular recently by leading to higher accuracy and dealing with a huge amount of data. Several research studies have been conducted using machine learning (ML) algorithms for flight delays prediction. Authors in [7] focused on predicting flight arrival time when aircraft are in the block. They considered features such as weather forecasts, time and airport congestion data using gradient boosting and linear regression algorithms. Authors in [8] conceived a predictive model to predict flightdelay using machine learning algorithms such as random forest, decision trees and multiple linear regression. As a result, Random Forest proved to be the best model. Authors in [9] applied machine learning techniques and statistical models in order to forecast flight delays in the United States. Supervised machine learning algorithms were selected by authors in [10] to produce a flight delay predictive system. They applied a sampling technique called costing which enhanced the performance of the cost-sensitive classifiers. Authors in [11] have implemented a machine learning approach to predict whether the flight is delayed or not using Gradient Boosting Classifier for the classification and Extra-Trees Regressor for the regression. In the research [12], flight and weather data have been utilized to predict on-time arrival traffic using machine learning methods. The proposed system revealed 77% of the accuracy for Random Forest. Wu et al. [13] adopted Support Vector Machine to estimate traffic departure delay using historical data of Beijing Capital International Airport. Compared to other algorithms, SVM was the best with an R^2 of 0.71 and a lower time of execution. Chakrabarty et al. [14] developed a system in which they estimate flight arrival delays operated by American Airlines utilizing four supervised machine learning algorithms. Authors in [15] picked features such us departure delay, origin airport, destination airport, distance between airports in order to predict flight arrival delays. The proposed model achieved an accuracy of 91% for the prediction. In order to solve flight-to-gate assignment problem, authors in [16] applied Random Forest for flight delay prediction. Authors in [17] performed causal data mining with machine learning algorithms. The proposed system achieved a high prediction accuracy of 91.97% for flight delays. Alla et al. [18] developed a multilayer perceptron neural network to forecast traffic arrival delays based on a selective training. The model with the selective training proved to be the best compared with the traditional training one with a higher accuracy and a lower computational time.

In the existing literature, authors have worked differently on the problem of flight delays. In this study, we propose a system that predicts air traffic delays employing supervised machine learning algorithms, where classification is performed to estimate the occurrence of the delay. New features which contribute to flight delays are considered for the first time in this study. Parameters optimization and data sampling were adopted in order to improve the accuracy of the proposed system.

3. PROPOSED METHODOLOGY

3.1 Problematic of the Research

Delay is regarded as the second important element in the air transportation system, behind safety. It is detrimental to passengers, airlines, and airport staff.

Numerous initiatives have been made to alleviate this problem. It is necessary to have a system that forecasts flight arrival delays and notifies airport staff, airlines, and travelers. Flight delays can be caused by a number of macro-level reasons, including poor weather, technical difficulties, late and unruly passengers, airport crowding, runway lines, a lack of airport infrastructure, delays with aircraft maintenance, delays with flight checklists, etc. In this study, our goal is to find and use new micro-level elements that contribute to flight delays in order to build a predictive model based on effective machine learning classifiers. Figure 1 outlines the steps used to produce and generate the proposed model. Flight on-time data are taken from the BTS database. Types of aircraft, constructors, and registration numbers are extracted from Github. Techniques for preprocessing and cleaning data are used to eliminate information that is duplicated, repetitive, noisy, and useless. We selected as features only necessary attributes which are pertinent and responsible for flight delays. Other micro-level elements of delays are explored in order to increase the effectiveness of the suggested system. We developed three novel features that as far as we know were the first attempt in this study: type of aircraft, arrival and departure periods of the day. We will prove later that the proposed new features have enhanced the accuracy of the suggested system. We split the data into 70% for training and 30% for testing. The dataset is imbalanced since the percentage of non delayed flights is higher than that of delayed flights. Therefore, sampling technique using a combination of Synthetic Minority Oversampling Technique (SMOTE) and Tomek Links is performed. For hyperparameters optimization, we chose Grid Search method. Finally, we evaluate the efficacy of the suggested model based on classification metrics, parameters tuning, data sampling and features selection.

3.2 Data Collection

First, data concerning on-time and delayed flights inside the US for all the year 2018 were taken from the BTS database. More than 100000 flight recording are available in our dataset. We chose this database due to its open access archives, trustworthiness, integrity and neutrality [19]. Following the rule established by the Federal Aviation Administration, we classified as delayed all the flights whose actual arrival times are 15 min or more bigger than their scheduled arrival times, and the other flights as on-time. Second, we took out data of aircraft constructor, types and registration from Github[1]. More than 400000 information about aircraft types, constructors and registrations are available in our dataset.

[1] https://github.com/RobAltenburg

3.3 Data Preprocessing

3.3.1 Data Cleaning

Database data often contain mistakes, omissions, misunderstandings and incompleteness. It could also be obtrusive and unreliable [20]. Therefore, data that can be understood and analyzed by computers and machine learning algorithms should be cleaned. We eliminated redundant information, empty fields, and missing values. We converted categorical data to numbers.

3.3.2 Data Balancing

The training data are imbalanced since the percentage of non delayed flights is 48% and that of delayed flights is only 14%, which makes data not equally presented. To solve the problem of class balancing, combined undersampling and oversampling techniques were used. The combination was chosen mainly because undersampling and oversampling, if utilized separately, have many problems. With a large sample of dataset, undersampling might be the reason of information loss. However, cases of overfitting and overgeneralization have been noticed when using oversampling methods [21]. Hence, the combination of both techniques is suggested to overcome this kind of issue.

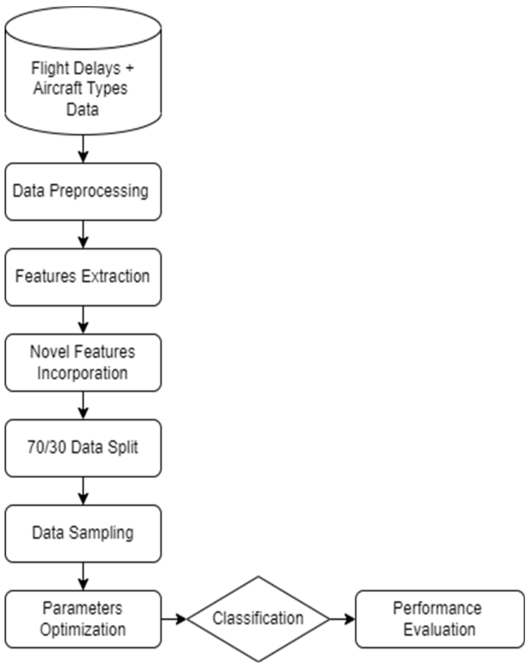

Figure 1. Proposed Methodology Process

Sampling algorithm used in this study is SMOTE-Tomek. The Synthetic Minority Oversampling Technique (SMOTE) is an oversampling method that generates synthetic minority class examples in the dataset. In order to over-sample the minority class, synthetic examples are introduced from a minor-class tuple and its nearest neighboring k-tuples [22]. SMOTE has been proven to be efficient with improving the accuracy of classifiers. According to [23], it is one of the most influential data preprocessing and sampling algorithms in machine learning and data mining. Tomek Technique is the ability to delete samples of data from the majority class which are close with the minority class data by Tomek links. SMOTE-Tomek technique is a combination of oversampling by SMOTE and undersampling by Tomek links. [24] used SMOTE-Tomek to improve the precision-recall of selling category in a beauty platform. Experimental results for a common diseases prediction conducted by authors in [25] showed that combining SMOTE with Tomek links technique is effective. Evaluation metrics and performance are evidently more improved and efficient than those of with only SMOTE.

3.4 Features Description

3.4.1 Features Extraction

Numerous variables and data fields were present in the extracted data. We removed non-essential features and kept just the ones that affect flight delays.

Airlines delay: According to BTS, airlines delays are brought on by technical problems of aircraft, pushback, icing/de-icing, fueling/refueling, charging/unloading, catering or cleaning operations. Besides, delay propagation, which can be generated from an earlier flight delay, is also considered as an important factor. Crew members, airport personnel stress, fatigue [19] or protestions [18], are also responsible for airline flight delays or cancellations.

Flight Distance: The travel time is influenced by the distance between the airports of Departure and Arrival. If it is greater, there is a larger chance that the flight may be delayed. According to [18], the delay increases with distance.

Departure Delay: Delays in departure lead to delays in arrival. The majority of delayed traffic at the destination airport, according to [26], derives from the outgoing airport. A delay occured at the origin will be propagated to the destination leading up to flight arrival delays [18].

3.4.2 Novel Features Incorporation

Type of Aircraft: Based on their performance, certain aircraft may be more adaptable in terms of being on time [27]. Flight delays are triggered by disparities in aircraft types, in accordance with [28]. The capacity of airports and the sky are generally defined by light aircraft such as Cessna, Piper, Beechcraft, etc. When a light aircraft leads the route, all the following flights will be delayed even if they are heavy or medium aircraft. That is because the velocity of the preceding aircraft is smaller. The following high-performance flights are not allowed then to increase the speed as long as the low-performance aircraft is preceding. This is how flight delays are generated.

The International Civil Aviation Organization (ICAO) [29] defines the longitudinal separation between two aircraft as follows:

- 10 minutes, if both aircraft have the same performance.
- 5 minutes if the preceding aircraft is maintaining a true airspeed of 37 km/h (20 kt) than the succeeding aircraft or higher.
- 3 minutes if the preceding aircraft is maintaining a true airspeed of 74 km/h (40 kt) than the succeeding aircraft or higher.

We deduce that the difference in performance determines the longitudinal separation between aircraft. If the preceding traffic is faster than the succeeding one, the separation is lower (since the second cannot catch the first one) and the risk of delay will be lower too. In Figure 2, we plot all the types of aircraft and their constructor used in our dataset organized by percentage of arrival delays. We notice, for example, that 27% of the flights performed by Cessna were delayed but only 13% by Boeing 739 were delayed too.

Part of the Day: Everyone is aware that the part of day is a major element to examine when traveling. Practically every traveler has a preferred boarding time. Some people like to travel early in the morning, while others prefer to take the flight late at night, and so on. There are also times of the day when there is a lot of traffic jams, which can cause traffic delays or cancellations. Due to visibility conditions, we expect more aircraft in the morning and afternoon than in the evening. To this end, we decided to add two novel aspects to our work: the departure and arrival time parts of the day, which represent the part-time of day when the flight departs and the part-time when the plane arrives, it is either in the early morning, late morning, afternoon, evening or night. To correlate flight delays of our model with the period of the day in which they were performed, we utilized data about Sunrise and Sunset times for the year 2018 in the United States from the world's top-ranking website[2] for date, time and time zones.

3.5 Machine Learning Classifiers

For the classification of a flight in two categories: Delayed or On-time, we employed machine learning

[2] World date, time and time zones, https://www.timeanddate.com

algorithms , e.g. K-Nearest-Neighbors, Decision Trees, Random Forest, and Multilayer Perceptron.

3.5.1 K-Nearest-Neighbors

K-NN algorithm is part of an Instance-Based Learning model in which a new instance is classified by making comparisons against the most similar and close instance in the training dataset. K is the number of similar instances and closest neighbors being compared with the new instance [30]. Using two populations designated by A and B, C is affected to the population A, if at least 1/2 k of the k values neighbors to C are originated from A.

3.5.2 Decision Trees

Decision trees are a popular and widely-used machine learning technique generally utilized to address prediction situations. Specifically, determining a discrete class labels (classification) or predicting a continuous value (regression) from a series of predictors [31]. It consists of nodes that constitute a rooted tree with no incoming edges. Every other node has only one incoming edge. An internal or test node is one that has outgoing edges. The other nodes are known as leaves (sometimes as terminal or decision nodes). Each internal node in a decision tree divides the instance space into two or more sub-spaces based on a discrete function of the input feature values. Each leaf is allocated to a class that represents the ideal target value. Additionally, the leaf might contain a probability vector showing the likelihood of the target attribute having a specific value. Based on the results of the tests along the path, instances are identified by routing them down from the tree's root to a leaf [32].

3.5.3 Random Forest

According to [33], Random Forests are an ensemble of classification and regression decision trees which constitute basic models that employ binary splits on predictors to produce forecasts. They were first proposed in 2001 by Breiman [34].
Segments of data are sampled, a randomized tree predictor is generated on each single piece, then the predictors are aggregated together [35]. Input features of the training dataset are used to randomly build the trees. Random forests generally outperform other classification models in terms of prediction accuracy. Qi Yanjun [36] used Random Forests in the area of Bioinformatics. Han et al. [37] employed it for route pavement maintenance, Belgiu et al. [38] for remote sensing, Farnaaz et al. [39] for network intrusion detection, Ali Khan [40] for modeling of surface water salinity, and so on.

3.5.4 Multilayer Perceptron

A Multilayer Perceptron (MLP) neural network has been considered an alternative to traditional statistical techniques [18] due to its ability to learn in a better way complex relationships between input and output patterns [41].

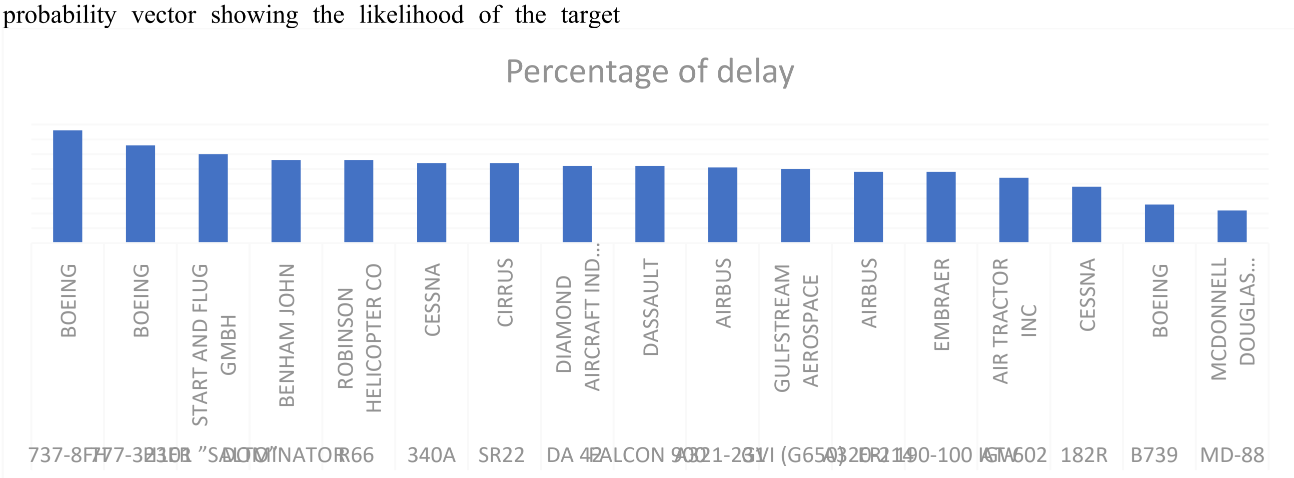

Figure 2. *Data Set Arrival delays by Type of Aircraft*

It is composed with an input layer, one or many hidden layers and an output layer. Each input has a connected weight, and each output has a transfer or activation function [42]. Equation (1) is used to calculate the output value of each neuron.

$$y_j = F(b + \sum_{i=1}^{n} x_i w_i) \qquad (1)$$

where x_i is the value of the input i, w_i is the weight, b is the bias, F is the activation function and y is the output value of the neuron j.
To find optimal parameters of our classifiers, improve accuracy, save time and energy, we adopted Grid Search

Table 1: Best Parameters using Grid Search technique

Algorithm	N°	Tuning Parameters	Score(10-fold)	Best Parameters
RandomForest	1	’maxfeatures’:[’sqrt’, ’log2’], ’n estimators’: [290, 300, 310]	95.24%	’max features’:’sqrt’, ’n estimators’: 300
RandomForest	2	’max features’:[’sqrt’, ’log2’], ’n estimators’: [250, 300, 350,400]	95.19%	’max features’:’sqrt’, ’n estimators’: 250
RandomForest	3	’max features’: [’sqrt’, ’log2’], ’n estimators’: [50, 100, 150,200]	95.20%	’max features’:’sqrt’, ’n estimators’: 200
RandomForest	4	’max features’: [’sqrt’, ’log2’], ’n estimators’: [275,300,325]	95.25%	’max features’:’sqrt’, ’n estimators’: 325
RandomForest	5	’max features’: [’sqrt’, ’log2’], ’n estimators’:[200,300,500,700]	95.27%	’max features’:’sqrt’, ’n estimators’: 700
RandomForest	6	’max features’:[’sqrt’, ’log2’], ’n estimators’: [500,1000,1500]	95.21%	’max features’:’sqrt’, ’n estimators’: 500
RandomForest	7	’max features’: [’sqrt’, ’log2’], ’n estimators’: [10,100,1000,2000],	95.23%	’max features’:’log2’, ’n estimators’: 2000
DecisionTrees	1	’criterion’: [’gini’, ’entropy’], ’max depth’: [5, 10, 15, 20, 25, 30, 35, 40, 45, 50], max leaf nodes’: [50, 100,150,200, 250, 300, 350, 400,450, 500]	93.18%	’criterion’:’entropy’, ’max_depth’:30, ’max leaf_nodes’:450
K-Nearest Neighbors	1	’n neighbors’: [5,10,15,20,25], ’p’: [2,1], ’metric’:[’str’,’callable’, ’minkowski]	90.50%	’metric’:’minkowski ’n_neighbors’:5, ’p’: 1
Multilayer perceptron	1	’hidden layer sizes’: [(5),(10), (15),(20)], ’activation’ :[’logistic’, ’tanh’, ’relu’], ’solver’ :[’lbfgs’,’sgd, ’adam’], ’alpha’ : [0.01,1e-6, 1e-2], ’learning rate’ : [’constant’, ’invscaling’,’adaptive’], ’max iter’ : [100,500,1000,2000]	98.37%	’activation’:’tanh’, ’alpha’: 0.01, ’hidden_layer_sizes’: (15,), ’learning_rate’:’constant ’max iter’: 1000, ’solver’: ’adam’
Multilayer perceptron	2	’hidden layer sizes’: [(10), (20),(30), (40),(50),(60),(70),(80), (90),(100)], Same other parameters above	98.42%	activation’:’tanh’, ’alpha’: 0.01, ’hidden_layer_sizes’:20,’learning rate’: ’adap-tive’, ’max_iter’:1000, ’solver’: ’adam’

method for parameters optimization using cross validation (cv=10). For the parameters optimization, we used the class GridSearchCV available in Scikit Learn [43] which is a package in Python. In order to tune the parameters, we followed the idea of [44] in which all possible combinations for Random Forest were considered to find optimal parameters. Table 1 provides the results of tuning parameters for Random Forest, Decision Trees and K-Nearest Neighbors. In Random Forest, 7 combinations of parameters were tested in order to extract the best one. 'max_features' represents the number of features to consider when looking for the best split and 'n_estimators' is the number of trees in the forest. In Decision Trees, 'criterion' contains the function used to measure the quality of a split, 'max_depth' is the maximum depth of the tree and 'max_leaf_nodes' represents the total number of terminal nodes in a tree. In K-Nearest Neighbors, the number of neighbors to use is interpreted by the parameter 'n_neighbors', the distance metric to use for the tree is performed by the parameter 'metric'. The power parameter for the Minkowski metric is 'p'. The parameters of each classifier are more explained in [43]. All potential combinations are considered and optimal ones are chosen. The best of Random Forest was number 5 with 95.27% of accuracy, 'max_features: sqrt' and 'n_estimators: 700' as the optimal parameters. Among the values selected for Decision Trees, the best were, 'max_leaf_nodes': 450, 'max_depth': 30, 'criterion': 'entropy', and 93.18% of accuracy. The best parameters in case of K-NN represents , 'n_neighbors': 5, 'p': 1, 'metric': 'minkowski' with 90.50% as the best accuracy.

To find optimal parameters of MLP classifier, we focused on prior studies analysis. According to [45], [46], [47], [48], [49], [50], [51], [52], [53], it is sufficient to use one single hidden layer to learn N arbitrary samples in a feed forward neural network with N hidden neurons when using almost any transfer function (Sigmoid, Logistic, etc.). But the network would then become very large with the augmentation of input samples. To reduce the number of hidden nodes, Tamura and Tateishi [54] proposed the use of a second hidden layer. Huang [55] demonstrated that in a two-hidden-layer network, the number of hidden nodes able to learn N samples with a small trivial error can be represented by Equation (2) and Equation (3).

$$L_1 = \sqrt{(m+2)N} + \frac{2\sqrt{N}}{\sqrt{m+2}} \qquad (2)$$

$$L_2 = \frac{m\sqrt{N}}{\sqrt{m+2}} \qquad (3)$$

While L_1 and L_2 are respectively the numbers of nodes in the first and second hidden layers and m the output neurons. Since L_1 and L_2 provide a very big number of nodes, the network will be more complex and may overfit the training data. According to [45], an accurate topology will contain fewer nodes than that suggested by both Equation 2 and Equation 3. Since no established methodology and exact solution was found to estimate an optimally or quasi-optimally ANN architecture, each researcher apply a different method to come up with satisfying results. Several methods are utilized such as trial and error, heuristic search, exhaustive search, Grid Search, Random Search, Bayesian, pruning and constructive algorithms. In our case, we apply Grid Search technique to find the best parameters. In the first selection of tuples of hidden layers and neurons, we were inspired from the study of Sonawane et al. [56] by using one single hidden layer and hidden neurons from 5 to 20 with a step of 5 ((5,) ,(10,), (15,), (20,)). The authors got the highest accuracy of 98.58 % for 20 neurons for Heart Disease Prediction. As a second selection, we adopted the structure used by Stefanovic et al. [57] to estimate Flight Time Deviation for Lithuanian Airports which consists of a Single layer with 10 to 100 hidden neurons by a step of 10 ((10,), (20,), (30,), (40,), (50,), (60,), (70,), (80,), (90,), (100,)). As a third run, we proposed a combination of single and two hidden layers ((50,), (100,), (50, 50), (100, 50), (100, 100)). For each run, we fixed the selection of activation function, solver, learning rate, number of iterations and regularization term (alpha) as represented in Table 1. The best parameters of our MLP classifier were: 'activation': 'tanh', 'alpha': 0.01, 'hidden layer sizes': (20,), 'learning rate': 'adaptive', 'max iteration': 1000, 'solver': 'adam' with a best accuracy of 98.42%.

4. EXPERIMENTAL RESULTS AND DISCUSSION

More than 100000 flight recording and 400000 aircraft types were collected before the study was carried out. Preprocessing techniques were performed in order to prepare data for the training. A percentage of 70\% was dedicated for the training set and 30\% for the testing. To investigate new aspects leading to flight delays and produce a good forecast, novel features were proposed in this paper. Efficient Machine learning classifiers were chosen in order to build a good predictive model. Parameters optimization and data balancing with respectively Grid Search and SMOTE-Tomek techniques were implemented.

4.1 Performance of Features

4.1.1 Part of the Day Effect

We discovered, once data based on novel features were studied and analysed, that departing aircraft are mostly behind schedule on nights, afternoons and late mornings. To explain more, for being less subject to congestion, morning is the most requested part of the day from passengers to travel in. Also, it has typically a moderate temperature and weather. Traffic jam, bad visibility, crosswinds, airport and airspace density, all cause delays especially in the afternoons and nights. Delays on afternoons and evenings can also be propagated to following flights which might extend the delay to cause late arrivals at night. Lastly, from Figure 3 and Figure 4, we observe that flights operated around midday were less influenced by delays.

4.1.2 Type of aircraft Effect

In Figure 5, we illustrate the results of the five most delayed types of aircraft. Arrival delays were particularly noticeable on Airbus and EMBRAER constructors. One suspected cause of the delay is the low performance compared, for example, with Boeing.

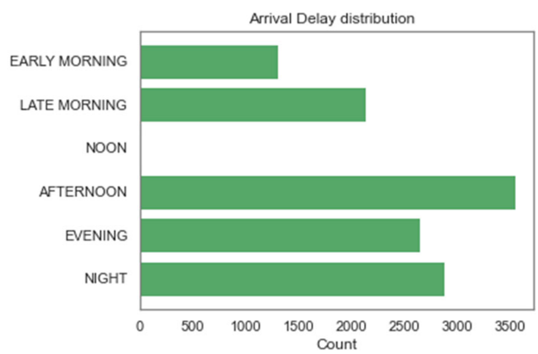

Figure 3: Delays on Arrival Against Period of the Day

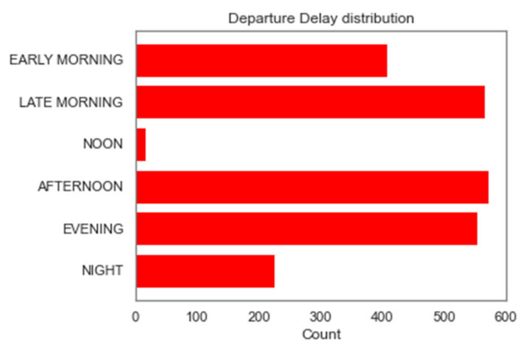

Figure 4: Delays on Departure Against Period of the Day

4.1.3 Airlines Effect

Figure 6 shows disparities within the delay occurred due to circumstances within the airline's control, for each company used in our dataset. We can explain that by the fact that each airline has its own policies about how to handle and avoid flight delays by defining Standard Operating Procedures (SOPs) or briefings in order to promote a culture of minimum delays. Any shortcomings in these procedures might contribute to delays.

4.1.4 Airport of Departure Effect

We decided to display in Figure 7 only the top-ten most delay-affected departure airports from our dataset. In this distribution, Los Angeles is the airport with the highest percentage of delays and San Juan is the airport with the lowest one. The difference in airports delays classification can be caused by many parameters such as: airports infrastructure, capacity, location, employees, operational management, etc.

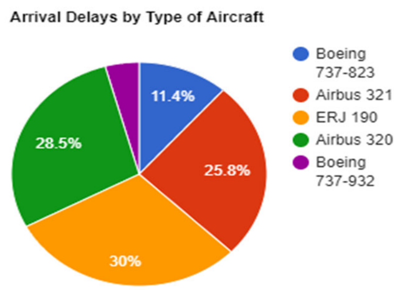

***Figure 5**: Arrival Delays Plots Based on Type of Aircraft*

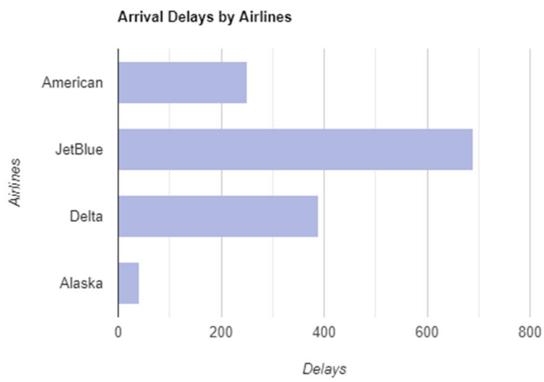

***Figure 6.** Arrival Delays Plots Based on Airlines*

4.2 Performance of Parameters Tuning

This work used the Grid Search approach to tune the parameters of Multilayer Perceptron, Random Forest, Decision Trees, and KNearest Neighbors. It has the objective of increasing the quality of the classification by identifying the best parameters. For that, it is primordial to study the evaluation results with and without parameters optimization. Based on Table 2 findings, we witness that all metrics of the four algorithms have improved properly.

4.3 Performance of Data Sampling

Once parameters tuning by Grid Search was done, we applied sampling technique using SMOTE-Tomek. From Table 3, the performance of all classifiers has increased when data balancing has been performed. We notice a slight difference between the accuracy and the recall which means that the model is able to better predict the positive values. The classification task is then handled properly.

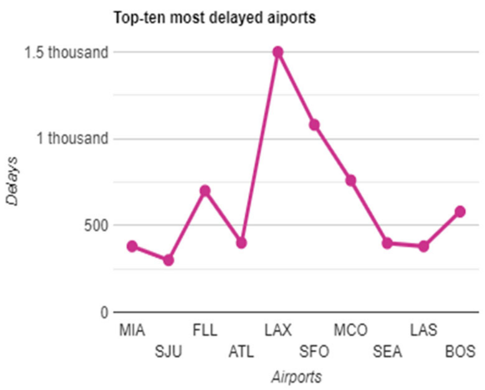

***Figure** 7: Top-ten Delayed Airports of Origin*

4.4 Confusion matrix

We determined confusion matrix measures for each classifier with sampling and parameters tuning. In Table 4, Table 5, Table 6 and Table 7, TP indicates the number of delayed flights that were correctly classified, and FP shows the ones which were wrongly assigned. Similarly, TN indicates the number of non-delayed (on-time) flights that were correctly classified, and FN shows the ones which were wrongly assigned.

The results show that parameters tuning and data sampling have successfully generated the best classifier which is Multilayer Perceptron with an accuracy of 98.42% and a higher number of correctly classified flights.

4. CONCLUSION

As air demand develops year after year, flight delay has become an essential hard study topic. For that, experts and scientists examined aircraft delays from many viewpoints.This research aimed to predict flight arrival delays. This research aimed to predict flight arrival delays. We designed and incorporated three additional novel features to improve the effectiveness of the proposed model: Departure Part of the day, Arrival Part of the day, and type of aircraft. So as to achieve optimal results, data sampling and parameters optimization were explored using SMOTE-Tomek and Grid Search tools. With Multilayer Perceptron, the model attained the greatest accuracy of 98.42%.

Table 2 : Classification performance with/without parameters tuning

Algorithm	Accuracy		Recall	
	Tuning	Without tuning	Tuning	Without tuning
MLP	97.44%	96.06%	97.66%	94.41%
RF	96.02%	93.89%	95.75%	77.76%
DT	92.34%	89.72%	89.46%	75.91%
KNN	91.88%	87.59%	92.75%	50.21%

Table 3 : Classification performance with/without data sampling

	Predicted NO (on-time)	Predicted YES (delayed)
Actual NO (on-time)	TN= 8968	FP= 172
Actual YES (delayed)	FN= 60	TP= 9170

Table 4: Confusion matrix of Multilayer Perceptron

	Predicted NO (on-time)	Predicted YES (delayed)
Actual NO (on-time)	TN= 8254	FP= 829
Actual YES (delayed)	FN= 649	TP= 8627

Table 5: Confusion matrix of KNN

	Predicted NO (on-time)	Predicted YES (delayed)
Actual NO (on-time)	TN= 8737	FP= 394
Actual YES (delayed)	FN= 384	TP= 8847

Table 6: Confusion matrix of Random Forest

	Predicted NO (on-time)	Predicted YES (delayed)
Actual NO (on-time)	TN= 8686	FP= 382
Actual YES (delayed)	FN= 860	TP= 8435

Table 7: Confusion matrix of Decision Trees

Algorithm	Accuracy		Recall	
	Sampling	Without sampling	Sampling	Without sampling
MLP	98.42%	96.65%	98.69%	87.13%
RF	96.01%	93.86%	96.07%	77.84%
DT	92.99%	92.51%	91.20%	73.71%
KNN	91.66%	87.63%	92.71%	48.44%

Airport authorities, companies, and customers can utilize the suggested model as a way to estimate aircraft arrival delays. It has also the potential to be used by air traffic control service providers in making decisions. In reality, if they are notified of flight arrival schedules and potential delays ahead of time, they will be advised of traffic rush hours in order to prepare for flight approach sequence in advance or call for reinforcement teams if it is obligatory.

The dataset only provides on-time statistics for non-stop national flights. In a future study, we plan to expand the dataset incorporating stopovers and foreign flights. Finally, it could be useful to integrate other attributes to the research in order to improve the classification of the model.

REFERENCES:

[1] Husni Idris, John-Paul Clarke, Rani Bhuva, and Laura Kang. Queuing model for taxi-out time estimation. Technical report, 2001.

[2] Yufeng Tu, 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(481):112–125, 2008.

[3] ZW Zhong, Dhir Varun, and YJ Lin. Studies for air traffic management r&d in the aseanregion context. Journal of Air Transport Management, 64:15–20, 2017.

[4] Kamila B Nogueira, Paulo HC Aguiar, and Li Weigang. Using ant algorithm to arrange taxiway sequencing in airport. International Journal of Computer Theory and Engineering, 6(4):357, 2014.

[5] Yu Jiang, Xinxing Xu, Honghai Zhang, and Yuxiao Luo. Taxiing route scheduling between taxiway and runway

in hub airport. Mathematical Problems in Engineering, 2015, 2015.

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

[7] Anna Achenbach 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.

[8] Anish M Kalliguddi and Aera K Leboulluec. Predictive modeling of aircraft flight delay. Universal Journal of Management, 5(10):485–491, 2017.

[9] Hemavathi.K Manisha.S Delhirani.S Bhuvaneshwari.R, Elakiya.S. Flight delay prediction using supervised machine learning. International Research Journal In Advanced Engineering And Technology (IRJAET), 5:4210–4216, 2019.

[10] Sun Choi, 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), pages 1–8. IEEE, 2017.

[11] Balasubramanian Thiagarajan, 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), pages 1–6. IEEE, 2017.

[12] Noriko Etani. 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(1):1–17, 2019.

[13] Weinan Wu, 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), pages 1–6. IEEE, 2019.

[14] Navoneel Chakrabarty, 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, pages 651–659. Springer, 2019.

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

[16] Zoutendijk, Micha, and Mihaela Mitici. "Probabilistic flight delay predictions using machine learning and applications to the flight-to-gate assignment problem." Aerospace 8, no. 6 (2021): 152.

[17] Truong, Dothang. "Using causal machine learning for predicting the risk of flight delays in air transportation." Journal of Air Transport Management 91 (2021): 101993.

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

[19] Hajar Alla, 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, pages 231–246. Springer, 2021a.

[20] Leonardo Moreira, 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), pages 1–8. IEEE, 2018.

[21] Haziqah Shamsudin, Umi Kalsom Yusof, Andal Jayalakshmi, and Mohd Nor Akmal Khalid. Combining oversampling and undersampling techniques for imbalanced classification: A comparative study using credit card fraudulent transaction dataset. In 2020 IEEE 16th International Conference on Control & Automation (ICCA), pages 803–808. IEEE, 2020.

[22] Sun Choi, Young Jin Kim, Simon Briceno, and Dimitri Mavris. Prediction of weather induced airline delays based on machine learning algorithms. In 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), pages 1–6. IEEE, 2016.

[23] Salvador Garc´ıa, Juli´an Luengo, and Francisco Herrera. Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. Knowledge-Based Systems, 98:1–29, 2016.

[24] Bern Jonathan, Panca Hadi Putra, and Yova Ruldeviyani. Observation imbalanced data text to predict users selling products on female daily with smote, tomek, and smotetomek. In 2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), pages 81–85. IEEE, 2020.

[25] Min Zeng, Beiji Zou, Faran Wei, Xiyao Liu, and Lei Wang. Effective prediction of three common diseases by combining smote with tomek links technique for imbalanced medical data. In 2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS), pages 225–228. IEEE, 2016.

[26] Sharma S. Sangoi H.-Raut R. Oza, S. and V.C Kotak. Flight delay prediction system using weighted multiple linear regression. International Journal of Engineering and Computer Science, 4:11668–11677, 2015.

[27] Maarten Tielrooij, 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, pages 577– 586, 2015.

[28] Aydan Cavcar and Mustafa Cavcar. Impact of aircraft performance characteristics on air traffic delays. Turkish Journal of Engineering and Environmental Sciences, 28(1):23–24, 2004.

[29] Doc-9854 International Civil Aviation Organization ICAO. Global Air Traffic Management Operational Concept, Doc 9854 AN/458. ICAO, 2005.

[30] Mladenović, Dušan, Slađana Janković, Stefan Zdravković, Snežana Mladenović, and Ana Uzelac. "Night Traffic Flow Prediction Using K-Nearest Neighbors Algorithm." Operational Research in Engineering Sciences: Theory and Applications 5, no. 1 (2022): 152-168.

[31] Talebi, Shawhin, John Waczak, Bharana A. Fernando, Arjun Sridhar, and David J. Lary. "Data-Driven EEG Band Discovery with Decision Trees." Sensors 22, no. 8 (2022): 3048.

[32] Rokach, Lior, and Oded Maimon. "Decision trees." In Data mining and knowledge discovery handbook, pp. 165-192. Springer, Boston, MA, 2005.

[33] Speiser, Jaime Lynn, Michael E. Miller, Janet Tooze, and Edward Ip. "A comparison of random forest variable selection methods for classification prediction modeling." Expert systems with applications 134 (2019): 93-101.

[34] Breiman, Leo. "Random forests." Machine learning 45, no. 1 (2001): 5-32.

[35] Biau, Gérard, and Erwan Scornet. "A random forest guided tour." Test 25, no. 2 (2016): 197-227.

[36] Qi, Yanjun. "Random forest for bioinformatics." In Ensemble machine learning, pp. 307-323. Springer, Boston, MA, 2012.

[37] Han, Chengjia, Tao Ma, Guangji Xu, Siyu Chen, and Ruoyun Huang. "Intelligent decision model of road maintenance based on improved weight random forest algorithm." International Journal of Pavement Engineering 23, no. 4 (2022): 985-997.

[38] Belgiu, Mariana, and Lucian Drăguţ. "Random forest in remote sensing: A review of applications and future directions." ISPRS journal of photogrammetry and remote sensing 114 (2016): 24-31.

[39] Farnaaz, Nabila, and M. A. Jabbar. "Random forest modeling for network intrusion detection system." Procedia Computer Science 89 (2016): 213-217.

[40] Khan, Mohsin Ali, M. Izhar Shah, Muhammad Faisal Javed, M. Ijaz Khan, Saim Rasheed, M. A. El-Shorbagy, Essam Roshdy El-Zahar, and M. Y. Malik. "Application of random forest for modelling of surface water salinity." Ain Shams Engineering Journal 13, no. 4 (2022): 101635.

[41] Muhammad Ehsan, R., Simon, S. P., Venkateswaran, P. R. (2017). Day-ahead forecasting of solar photovoltaic output power using multilayer perceptron. Neural Computing and Applications, 28(12), 3981-3992.

[42] Soares, F. A. A., Flores, E. L., Cabacinha, C. D., Carrijo, G. A., Veiga, A. C. P. (2013). Recursive diameter prediction for calculating merchantable volume of eucalyptus clones using Multilayer Perceptron. Neural Computing and Applications, 22(7), 1407-1418.

[43] Scikit-Learn, (2021) . Metrics and scoring quantifying the quality of predictions. http://scikit-learn.org/stable/modulesmodelevaluation.html.Accessed : 2021−07−26.

[44] Muhammad Murtadha Ramadhan, 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, (cece), 2017.

[45] Ramadhan, M. M., Sitanggang, I. S., Nasution, F. R., Ghifari, A. (2017). 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.

[46] Stathakis, D. (2009). How many hidden layers and nodes?. International Journal of Remote Sensing, 30(8), 2133-2147.

[47] Coulibaly, P., Anctil, F., Bobée, B. (1999). Prévision hydrologique par réseaux de neurones artificiels: état de l’art. Canadian Journal of civil engineering, 26(3), 293-304.

[48] Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. Mathematics of control, signals and systems, 2(4), 303-314.

[49] Hornik, K., Stinchcombe, M., White, H. (1989). Multilayer feedforward networks are universal approximators. Neural networks, 2(5), 359-366.

[50] Özgür, K. İ. Ş. İ. "Daily river flow forecasting using artificial neural networks and auto-regressive models." Turkish Journal of Engineering and Environmental Sciences 29, no. 1 (2005): 9-20.

[51] Rezaeian Zadeh, M., Amin, S., Khalili, D., Singh, V. P. (2010). Daily outflow prediction by multi layer perceptron with logistic sigmoid and tangent sigmoid activation functions. Water resources management, 24(11), 2673-2688.

[52] Rezaeian-Zadeh, M., Tabari, H., Abghari, H. (2013). Prediction of monthly discharge volume by different artificial neural network algorithms in semi-arid regions. Arabian Journal of Geosciences, 6(7), 2529-2537.

[53] Zhang, G., Patuwo, B. E., Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. International journal of forecasting, 14(1), 35-62.

[54] Tamura, S. I., Tateishi, M. (1997). Capabilities of a four-layered feedforward neural network: four layers versus three. IEEE Transactions on Neural Networks, 8(2), 251-255.

[55] Huang, G. B. (2003). Learning capability and storage capacity of two-hidden-layer feedforward networks. IEEE transactions on neural networks, 14(2), 274-281.

[56] Sonawane, J. S., Patil, D. R. (2014, February). Prediction of heart disease using multilayer perceptron neural network. In International conference on information communication and embedded systems (ICICES2014) (pp. 1-6). IEEE.

[57] Stefanovič, Pavel, Rokas Štrimaitis, and Olga Kurasova. "Prediction of flight time deviation for lithuanian airports using supervised machine learning model." Computational intelligence and neuroscience 2020 (2020).