MEF UNIVERSITY

FLIGHT DELAY PREDICTION

Capstone Project

Mustafa Kurt

İSTANBUL, 2019

MEF UNIVERSITY

FLIGHT DELAY PREDICTION

Capstone Project

Mustafa Kurt

Advisor: Asst. Prof. Duygu TAŞ KÜTEN

İSTANBUL, 2019

MEF UNIVERSITY

Name of the project: FLIGHT DELAY PREDICTION
Name/Last Name of the Student: Mustafa Kurt
Date of Thesis Defense: /09/2019

I hereby state that the graduation project prepared by Mustafa Kurt has been completed under my supervision. I accept this work as a "Graduation Project".

/09/2019

Asst. Prof. Duygu TAŞ KÜTEN

I hereby state that I have examined this graduation project by Mustafa Kurt which is accepted by his supervisor. This work is acceptable as a graduation project and the student is eligible to take the graduation project examination.

/09/2019

Director
of
Big Data Analytics Program

We hereby state that we have held the graduation examination of _____ and agree that the student has satisfied all requirements.

THE EXAMINATION COMMITTEE

Committee Member	Signature
1. Asst. Prof. Duygu TAŞ KÜTEN	……..
2. Prof. Dr. Özgür ÖZLÜK	……..

Academic Honesty Pledge

I promise not to collaborate with anyone, not to seek or accept any outside help, and not to give any help to others.

I understand that all resources in print or on the web must be explicitly cited.

In keeping with MEF University's ideals, I pledge that this work is my own and that I have neither given nor received inappropriate assistance in preparing it.

Name	Date	Signature

EXECUTIVE SUMMARY

FLIGHT DELAY PREDICTION

Mustafa Kurt

Advisor: Asst. Prof. Duygu TAŞ KÜTEN

SEPTEMBER, 2019, 16 pages

This study aims to create a model to predict flight departure delays. Various factors might affect a flight delay, and thus different features might be selected as input to create a model concerning priorities and the power of control over the features for the party who makes the analysis.

In this study, domestic commercial flights in the U.S. operated in August 2018 are studied. Besides, airplane, passenger boarding, and cargo data are combined with flight data to benefit from possible insights related to these factors.

For predicting the flight delays, machine learning methods such as decision trees, random forest, bagging classifier, extra trees classifier, gradient boosting and xgboost classifier are used and results are analyzed.

Further studies could be adding extra features such as data related to flight planning, personnel data, loading data, data about technical processes to prepare a plane to a flight to improve prediction capacity.

Key Words: Flight Delays, Prediction, Machine Learning, Classification, Tree-Based Algorithms

ÖZET

UÇUŞ GECİKME TAHMİNLEMESİ

Mustafa Kurt

Dr. Öğretim Üyesi Duygu TAŞ KÜTEN

AĞUSTOS, 2019, 16 Sayfa

Bu çalışma, uçuş kalkış gecikmelerini tahmin etmek için bir model oluşturmayı amaçlamaktadır. Çeşitli faktörler bir uçuş gecikmesini etkileyebilir ve bu nedenle, öncelikleri ve analizi yapan tarafın faktörler üzerindeki kontrolüne göre bir model oluşturmak için girdi olarak farklı özellikler seçilebilir.

Bu çalışmada, ABD'de Ağustos 2018'de düzenlenen iç hat uçuşları incelenmiştir. Ayrıca, uçak, yolcu uçağı ve kargo verileri, bu faktörlerle ilgili olası iç görülerden yararlanmak için uçuş verileriyle birleştirilmiştir.

Uçuş gecikmelerini tahmin etmek için karar ağacı, rastgele orman, torbalama sınıflandırcı, ekstra ağaçlar, grade takviyeli sınflandırıcı ve ekstra grade takviyeli sınıflandırıcı gibi makine öğrenme metotları ve sonuçları analiz edilmiştir.

Çalışmanın ileriki aşamaları için uçuş planlama verileri, personel verileri, yükleme verileri ve bir uçağı uçuşa hazırlamak için teknik süreçler ile ilgili veriler kullanılarak modelin tahminleme kapasitesi artırılabilir.

Anahtar Kelimeler: Uçuş Gecikmeleri, Tahminleme, Makine Öğrenmesi, Sınıflandırma, Karar ağacı temelli algoritmalar

TABLE OF CONTENTS

.

1. INTRODUCTION

As the number of flights grows every day, the flight networks become more complex. This helps us access more destinations and get things done in different locations in shorter amounts of time. Handling all these operations can sometimes become hard and this brings the risk of delay of a flight. Flight delays can lead to some negative effects on passengers, airlines, airline employees, economy and also the environment.

This study aims to create a model to predict flight departure delays. Various factors related to features of the airplane, properties of the airport, number of the passengers and the amount of cargo, the selected processes of preparing the airplane, the weather conditions, the previous flights of the airplane, the operational intensity of the airline and the airport might affect a flight delay and more factors can be added into this list. So, different features might be selected as an input to create a model concerning priorities and the level of control on factors by the party who makes the analysis.

In this study, domestic commercial flights in the U.S. in August 2018 are studied. Also, airplane, passenger boarding, and cargo data are combined with flight data to benefit from possible insights related to these factors. All data are gathered from the Bureau of Transportation Statistics web pages.

For predicting the flight delays supervised machine learning methods such as decision trees, random forest, bagging classifier, extra trees classifier, gradient boosting and xgboost classifier used and obtained results are analyzed.

1.1. Flight Delay Prediction: Literature Survey

Chen (2019) criticized previous studies about flight delay predictions that they produce technically correct results but they lack explanations of reasons. Holden (2019) indicated that Google's flight delay prediction product shares delay information when they are 85% confident, and they also provide the reason for the delay. Martinez (2012) predicted flight arrival delays with a long-term understanding which focuses on factors not changing daily. In this study, the author used a data-driven method considering past observations and concluded that more data do not improve global models because of the improvements in traffic management systems and airline's processes. The reason for this is if the time range of data increased it would cause comparing flights that don't use the same technology or processes. Sternberg et al. (2017) presented a study covering a wide range of

studies on flight delay predictions. The studies in the field grouped into three subjects, which are delay propagation, root delay, and cancellations. Features used in studies are grouped such as features related to planning, temporal reasons, weather, spatial, operations and state of the system. In the data management step, studies followed cleaning, feature selection, data transformation, outlier removal, correlation, normalization and discretization tasks. Some of the machine learning algorithms implemented in these studies are K-nearest neighbors, neural networks, support vector machines, random forest and boosting classifiers. Zonglei et al. (2009) combined supervised and unsupervised methods. Khaksar and Sheikholeslami (n.d.) explained the context in two parts, (i) a planning phase including flight scheduling, fleet assignment, fleet routing, crew assignment and (ii) operational phase including revenue management and gate assignment. They compared bayesian, *k*-means, decision tree, cluster classification, random forest, and hybrid methodologies. The authors obtained the highest score from a hybrid model of decision tree and *k*-means classification methods.

1.2. Tree-Based Algorithms: Literature Survey

A decision tree is the root of tree-based algorithms. Decision trees are created by dividing the data into groups with decision rules by regarding the values of features. Navlani (2018) explains how it works in three steps. First, the best attribute is selected to divide the data. Then, data is divided into subsets according to the selected attribute. Finally, this process is repeated until no more observations or features are left or all of the remaining observation values belong to the same class. However, decision trees carry the risk of overfitting. Moreover, random forests use random samples of data with the same distributions and create many decision trees. Each of these trees decides for the class and the most popular decision is selected as a result (Breiman, 2001). Since samples are random and the number of trees is many, models using random forest do not overfit. Boosting algorithms use many weak learning models based on decision trees and create a strong model by combining these models. Each time, a tree is added to the model and weights of features in that model are fixed and the next model is created for not well-classified observations (Nelson, 2019).

1.3. Sections

The remaining of the report is organized as follows. In Section 2, datasets, features, the source of the data, and the data pre-processing steps are explained. Section 3 defines the objective of the study and the project scope. Moreover, the methods, tools, and techniques benefited are listed in this section. In Section 4, the obtained results are analyzed and different methods are compared. Finally comments about further studies, value delivered, potential social and the environmental impact of the study are presented in Section 5.

2. ABOUT THE DATA

The data are gathered from the information published by the Bureau of Transportation Statistics (Bureau of Transportation, 2018) and by the Federal Aviation Administration (Federal Aviation Administration, 2019). The following data sources are used together: (i) “On-Time: Marketing Carrier On-Time Performance” is referred to as the flight data, (ii) “T-100 Domestic Segment (U.S. Carriers)” is referred to as the passenger and cargo data, and (iii) “Releasable Aircraft Database” is referred to as the Aircraft data.

Flight data is one of the open data sources that researchers refer to flight-related studies. This data includes monthly reports of each airline company with the information about time, airline, origin, destination, departure performance, arrival performance, cancellation and diversion information, flight summaries, cause of delay (if any) and diverted airport information of each flight (if diversion occurred). Features included in this study are explained in Section 2.1.

Passenger data is also reported monthly. It has carrier, origin, destination, aircraft, time, class and summary information of domestic segment flights. The number of passengers transported field from this data is included in the study with the matching carrier, origin and destination combination.

Aircraft data has information about the registered aircraft and is added into this project by combining the tail number of aircraft. The aim of using this data is to gain information about whether aircraft type or age has a notable effect on flight delays.

2.1. Features

Features that are selected from the flight data and employed in this project can be listed as follows:

- Day of Month: The day number of the month (e.g. 1,2,3…31)
- Day of Week: The weekday of flight (e.g. Monday, Tuesday)
- Operated or Branded Code Share Partners: Operated by a single company or by more than one company.
- Operating Airline: The airline company operated flight (e.g. AA, UA)
- Tail Number: Unique number assigned to each plane
- Origin Airport ID: The origin airport of the flight

- Destination Airport ID: The destination airport of the flight
- CRS Departure Time: Planned Departure Time
- Departure Time Blk: Hourly time interval of the flights (e.g. 10.00-11:00)
- Departure Delay Minutes: The departure delay of the flight in minutes.
- Distance: Distance of origin and destination airports in miles.

In addition to these features, additional features are created to check if the volume of passengers affects the departure performance:

- Flight order: For each aircraft, the number of flights operated in a day is counted and the order of each flight is known. This feature is used to check if the number of flights by an aircraft affects the departure performance.
- Airport flight order: For each airport, the number of flights in an hour is counted. This feature is used to check if the number of flights from an airport affects the departure performance for any specific time interval.
- Carrier flight order: For each operating airline the number of flights in an hour is counted. This feature is used to check if the number of flights by an airline affects the departure performance for any specific time interval.
- Selected features from Passenger and cargo data are:
- Passengers: The number of passengers transported

Finally, the age of the aircraft is included as a feature to check whether factors related to aircraft's age affect the departure performance.

2.2. Exploratory Data Analysis

After removing the NA values and outliers, the final data set had 638,776 rows and 18 columns. In the final data set, 38.29% of the flights are labeled as delayed.

Figure 1 presents the percentage of delayed flights with respect to the day of the month. As can be seen, almost all days of the last two weeks have a delay percentage of less than average. Such a relationship might be even more meaningful in case the delay performance is analyzed over the years.

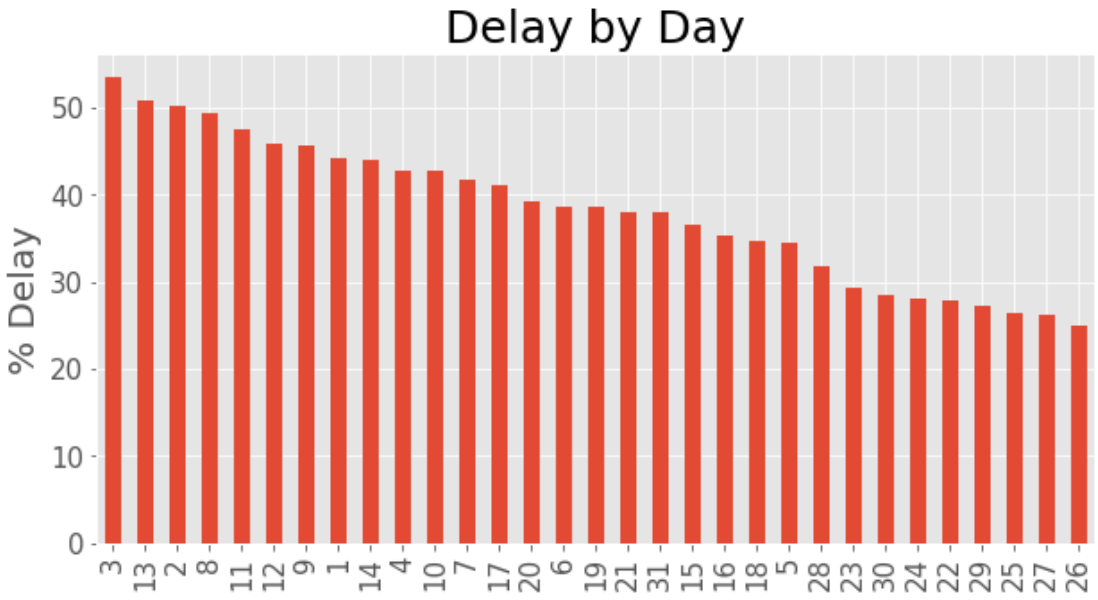

Figure 1: The percentage of delayed flights with respect to the day of the month

Figure 2 shows the percentage of delayed flights with respect to the day of the week. A similar observation is obtained: values are similar over the days of the week. So, this feature is not expected to be a good one for classification.

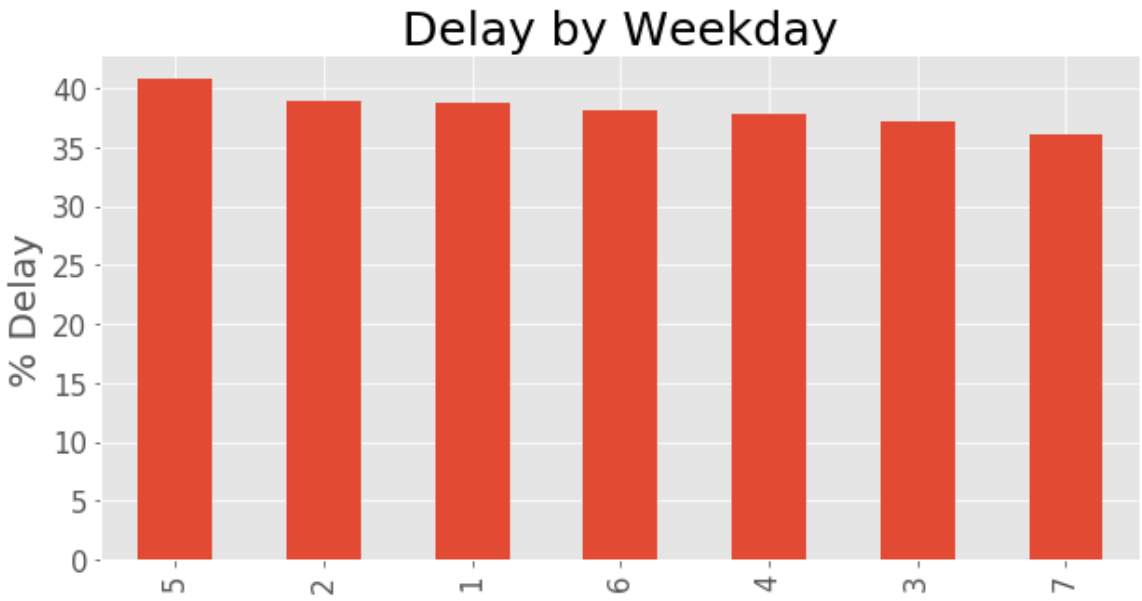

Figure 2: The percentage of delayed flights with respect to the day of week

When the level of focus is increased to hourly time intervals of flights (see Figure 3), it can be said that delays observed at flights operated before 1 pm are less than averages. Moreover, most delays occur between 5 pm and 11 pm, highest at [6 pm – 7 pm).

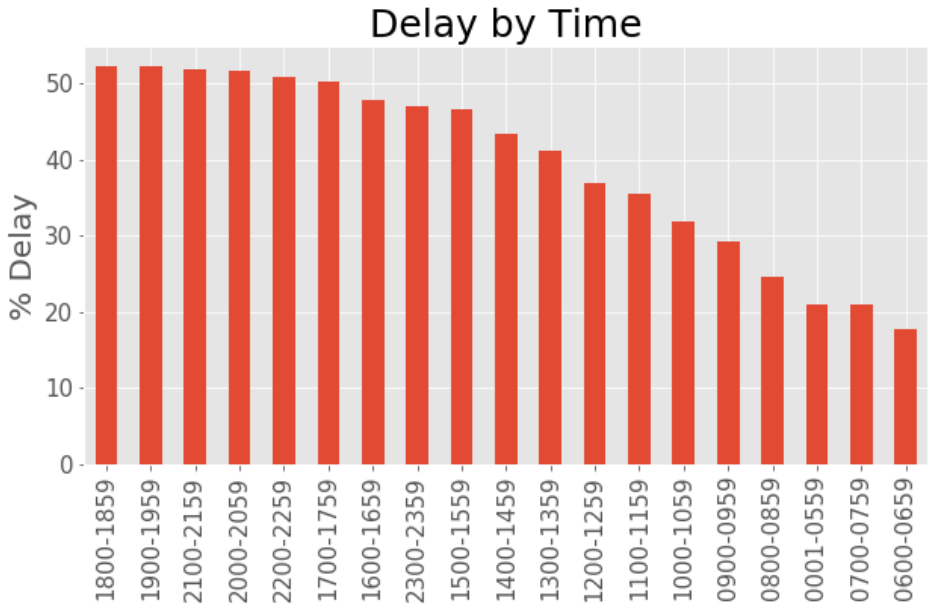

Figure 3: The percentage of delayed flights with respect to departure time

For the analysis of departure airport, the first 50 airports that have the highest is selected. Figure 4 represents the percentage of delayed flights with respect to these selected airports. This figure shows that there is no clear relationship between the delayed flights and the departure airports.

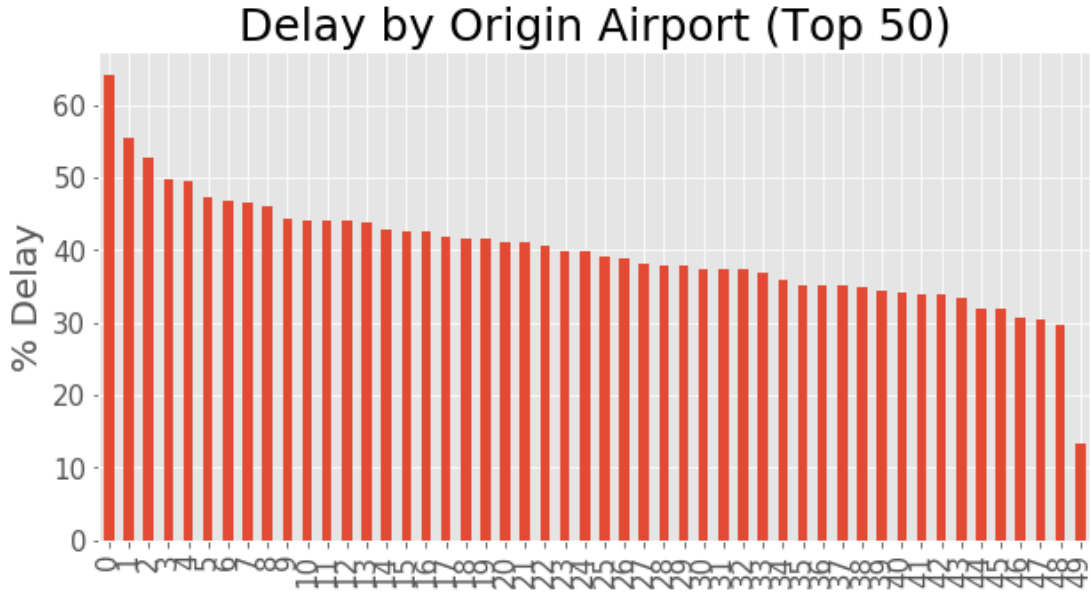

Figure 4: The percentage of delayed flights with respect to the departure airport

Figure 5 shows the percentage of delayed flights handled by a single airline is higher than co-operated flights.

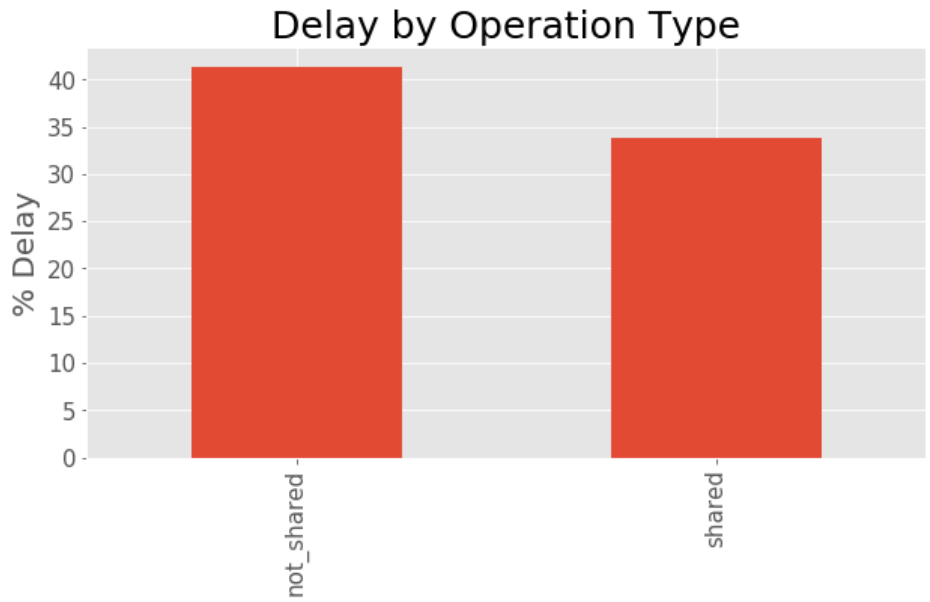

Figure 5: The percentage of delayed flights with respect to the operation type

Figure 6 represents the percentages of delayed flights with respect to the age of the airplane. As it is easily observed, values are similar over the age of the airplane. Therefore, it can be concluded that there is no relationship between the age of the airplane and the delay possibility.

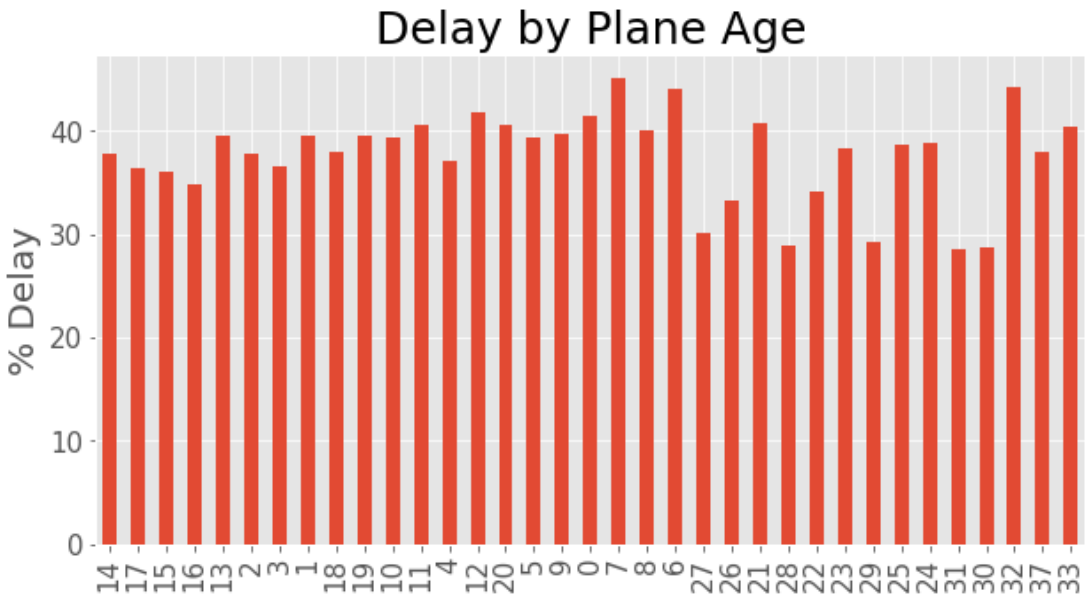

Figure 6: The percentage of delayed flights with respect to the age of the airplane

3. PROJECT DEFINITION

In this section, the problem statement, project objectives and scope are defined. Then methodologies, tools, and techniques are discussed.

3.1. Problem Statement

From the thousands of flights operated every day, some of them start with a delay. A flight delay might affect many stakeholders such as passengers, airlines, airline workers, airports, airport ground operation workers, airline service suppliers, air traffic management units and many others. On the other hand, all of these stakeholders also contribute to specific factors such as passenger boarding, flight and crew planning, loading fuel, queuing planes to the runway, when combined may result in a flight delay. And one might assume preventing flight delays necessitates effort from all related parties. So, working on a high-level model, pointing areas of development for each stakeholder and then building models for each part of stakeholders' actions might create a push for all related parties.

3.2. Project Objectives

The objective of the project is to develop machine learning models to predict flight delays. Then, a model is selected, where its performance is increased by implementing and making necessary hyperparameter tuning steps.

3.3. Project Scope

The delay referred to in this study is departure delay, which is studied in a long-term perspective. In other words, features showing short term effects such as arrival time of flights are not included in models. Otherwise, we might expect if a plane arrived late from a previous flight it would have a higher potential of departure delay on its next flight. So, we might have a technically correct model but we would not be explaining reasons for delays. Besides, both in terms of weather and occupation, seasonality might have an important effect on delay performance that can be a topic of another study. To overcome the seasonality effect August 2018 data is selected for this study.

3.4. Methods, Tools, and Techniques

Earlier phases of selecting data, preliminary work to explore details of data are done with R. After that, pre-processing steps are completed using Anaconda Spyder. Final wrap up, modeling and hyperparameter tuning operations are handled on Google Colab.

In this study, three different data sets are combined to form a final dataset. The plane data is merged with the flight data using tail number column at flight data and n-number column which corresponds to tail numbers of planes at plane data. Originally some of the tail number rows started with “N”, so “N” values are dropped using a lambda function that applies a defined operation to all rows. The unique carrier id, origin airport id, and destination airport id columns are used in both datasets to be able to include the passenger data. For outlier treatment step, values with distance more than 1.5 IQR identified as an outlier and removed.

Three new features are created from the data to benefit from possible insights. It is expected that as the number of flights that a plane operates increases in a day, the possibility of delay might increase. So, the first feature created is daily flight order. This feature is created by first sorting the data by the day of the month, the tail number and the CRS departure time. After this ordering, 1 is assigned to the first flight as daily flight order and this number is increased 1 for each next flight by CRS departure time that has the same day of the month and tail number information. It is expected that as the number of flights that a carrier operates increases from an airport in hourly time intervals, the workload might increase the delay possibility. Therefore, the second feature, which is the carrier flight order, is created by sorting the data by day of the month, unique carrier, origin airport ID and departure time block. After this ordering, 1 is assigned to the first flight as carrier flight order and this number is increased 1 for each next flight by departure time block that has the same day of the month, unique carrier and origin airport id information. Finally, the airport flight order is created by sorting the data with respect to the day of the month, origin airport ID and the departure time interval. After this ordering, 1 is assigned to the first flight as airport flight order and this number is increased 1 for each next flight by departure time block with the same day of month and origin airport id information with an expectation of including airport traffic.

Additionally, operated or branded codeshare partners feature is showing specific airlines included in code share operation. This feature is mutated to show only whether a

flight is operated by only an airline or more than one airline so the airline included in code share operation is not listed for simplification.

After preparing the data for modeling, decision trees, random forest, bagging classifier, extra trees classifier, gradient boosting and xgboost classifier models are fitted with default settings. Table 1 shows the accuracy, recall and F1 scores of models with the default setting.

Table 1: The accuracy, Recall and F1 scores of models with the default setting and 10 as max_depth

SCORE NAME	MODEL	SCORE
ACCURACY SCORES	Decision Tree	0.6777
	Random Forest	0.6808
	Extra Trees	0.6686
	Bagging	0.6867
	Gradient Boosting	0.7172
	XGBoosting	0.7165
RECALL SCORES	Decision Tree	0.3854
	Random Forest	0.3939
	Extra Trees	0.3099
	Bagging	0.4091
	Gradient Boosting	0.4971
	XGBoosting	0.4934
F1 SCORES	Decision Tree	0.4783
	Random Forest	0.4862
	Extra Trees	0.4176
	Bagging	0.5002
	Gradient Boosting	0.5740
	XGBoosting	0.5716

After running models with default settings, feature importance is checked, and mail and freight features are dropped before further improvements. In the following steps, features included in the models are changed iteratively by excluding one feature each time. Moreover, newly created features, such as daily flight order, carrier flight order and airport flight order, are also included in the model with modified versions. At created versions, the order is increased by 1, for the modified versions order is increased by 3 for carrier flight order, this means every 3 flight is grouped into 1, likewise every 15 flights from an airport

grouped into 1 for airport flight order. By doing so, it is aimed to create groups for carriers and airports to simplify features.

Also, origin airports are grouped with respect to the number of flights operated from with thresholds like the top 20 to the top 50 airports. After conducting several experiments with different versions of features, and using multiple combinations of features on default settings of models, accuracy and F1 scores are checked for further improvements. Finally, decision tree, gradient boosting and xgboost classifier methods are chosen for hyperparameter tuning steps.

Mohtadi (2017) suggests a method for gradient boosting hyperparameter tuning. In this method, model settings are checked one by one, and AUC scores of train and test datasets are plotted. After that, results are checked to decide the optimal setting. In this project learning rate, *n*-estimators, max depth, min-samples split, min-sample leaf, and max feature settings are iteratively checked by AUC scores of training and test data sets and values for these settings are decided according to the changes in AUC curve where AUC scores of training and test data sets show significant change..

For decision tree classifier, hyperparameter tuning max depth, min samples split, min samples leaf, and max feature settings are iteratively checked by AUC scores of training and test data sets.

For xgboost classifier, hyperparameter tuning learning rate, n-estimators, max depth, gamma, sub sample, col sample by tree settings are iteratively checked by accuracy and F1 scores of training and test data sets.

4. RESULTS

4.1. Overview

As a result, decision tree, random forest, and gradient boosting models achieved very similar scores according to accuracy and F1. Xgboost classifier performed slightly better and tuning hyperparameters did not bring a dramatic increase in these scores.

A high number of dimensions in categorical features might result in high variance like observed in origin airport IDs. Grouping categorical variables to decrease variance added more value than using parameter tuning and contributed to decreased computation time.

Also, features that are originally not included in data, such as the age of the plane, daily flight order, carrier flight order, airport flight order, turned out quite effective to explain results.

4.2. Analysis Expected vs Obtained

At the beginning of the study and before the literature survey, it was expected to find clear relations between features and flight delays. However, the complexity of the problem necessitates broader datasets to completely explain processes. There are many sub-processes to prepare a plane to the flight. For example, an airline removed seatback pockets to cut boarding times and managed to succeed (Elliot, 2019). Also, airlines try different methods for boarding passengers and even the size of the hand luggage's effect boarding (Barro, 2019). In addition, (i) loading cargo, food and fuel, (ii) boarding priority customers (such as frequent flyers, CIP's, VIP's) or customers requiring special assistance, (iii) managing overbook, (iv) completing security checks, taxi operations, (v) the distances between gates and pitches and (vi) ground traffic can also be included in the analyses to obtain more accurate results.

Additionally, airports were expected to play a more significant role in determining the delay possibility. However, the departure airports of flights were not on higher ranks in terms of feature importance. Unlike airports, the age of the airplane became an important feature to explain whether a flight is delayed or not.

5. SOCIAL AND ETHICAL ASPECTS

This study is carried by a long-term prediction approach that can be implemented long before a flight's departure. So, features as a result of events occurring close to the flight time, as the arrival time of the aircraft, did not include in the models.

Also, while similar studies in flight delay prediction tend to benefit from data related to flights, in this study it is attempted to enrich the models with data related to aircraft and passengers and cargo. Hopefully, this study with other increasing numbers of studies in this subject contributes to improving on-time performances of flights so that both airline customers, airline personnel, airlines itself can benefit from time left to themselves saved from delays. Also shrinking extra fuel consumption would be important for the environment. Finally rebound of opportunity costs of delays can add value to the economy.

6. REFERENCES

[1] Bureau of Transportation Statistics. (2018). *On-Time: Marketing Carrier On-Time Performance* Retrieved from https://www.transtats.bts.gov/Tables.asp?DB_ID=120&DB_Name=Airline%20On-Time%20Performance%20Data&DB_Short_Name=On-Time

[2] Bureau of Transportation Statistics. (2018). *T-100 Domestic Segment (U.S. Carriers)* Retrieved from https://www.transtats.bts.gov/Tables.asp?DB_ID=110&DB_Name=Air%20Carrier%20Statistics%20%28Form%2041%20Traffic%29-%20%20U.S.%20Carriers&DB_Short_Name=Air%20Carriers#

[3] Federal Aviation Administration. (2019). *Releasable Aircraft Database* Retrieved from https://www.faa.gov/licenses_certificates/aircraft_certification/aircraft_registry/releasable_aircraft_download/

[4] Elliot, A. F. (2019). How long does it take to turn a plane around – and what's the fastest way to board? *The Telegraph.* Retrieved from https://www.telegraph.co.uk/travel/travel-truths/plane-turnaround-procedures/

[5] Barro, J. (2019). Here's Why Airplane Boarding Got So Ridiculous. *The New York Magazine Inelligencer* Retrieved from http://nymag.com/intelligencer/2019/05/heres-why-airplane-boarding-got-so-ridiculous.html

[6] Chen, Stephen (2019). *Failing to Land Flight Delay Predictions https://towardsdatascience.com/failing-to-land-flight-delay-predictions-a281689dd602*

[7] Holden, Richard (2018). *For the holidays and beyond, your travel planning guide is here* https://www.blog.google/products/flights-hotels/travel-planning-guide-for-the-holidays-and-beyond/

[8] Martinez Vincent (2012). *Flight Delay Prediction* (Master's Thesis). Available from https://www.semanticscholar.org/paper/Flight-Delay-Prediction-Master-Thesis-Martinez/84daebe3efe5cdc9678d791d4897cd16fc197d92

[9] Sternberg, Alice & Soares, Jorge & Carvalho, Diego & Ogasawara, Eduardo. (2017). *A Review on Flight Delay Prediction*.

[10] Zonglei, Lu & Jiandong, Wang & Guansheng, Zheng. (2009). *A New Method to Alarm Large Scale of Flights Delay Based on Machine Learning.* 589-592. 10.1109/KAM.2008.18

[11] Khaksar, Hassan & Sheikholeslami, Abdolrreza(n.d.). *Airline Delay Prediction by Machine Learning Algorithms.* http://scientiairanica.sharif.edu/article_20020_d43b2e5c29cb07fcb651da6bd2005d30.pdf

[12] Mohtadi Ben Fraj (2017 December 24). *In-Depth: Parameter tuning for Gradient Boosting* https://medium.com/all-things-ai/in-depth-parameter-tuning-for-gradient-boosting-3363992e9bae

[13] Avinash, Navlani (2018 December 28). Decision Tree Classification in Python. Retrieved from https://www.datacamp.com/community/tutorials/decision-tree-classification-python

[14] Breiman, L. Machine Learning (2001) 45: 5. https://doi.org/10.1023/A:1010933404324

[15] Dan, Nelson (2019 July 17) Gradient Boosting Classifiers in Python with Scikit-Learn Retrieved from https://stackabuse.com/gradient-boosting-classifiers-in-python-with-scikit-learn/