Ministry of Education and Science of the Republic of Kazakhstan
Suleyman Demirel University

Yerkebulan Zhumagali

Weather-induced flight delay prediction using Machine Learning

THESIS

Presented in Partial Fulfillment for the
Degree of Master of Science in Computer Science
(degree code: 7M061002)

Department of Computer Science
Faculty of Engineering and Natural Sciences

Supervisor: **PhD, Andrey Bogdanchikov**

Kaskelen, 2022

Suleyman Demirel University

Faculty of Engineering and Natural Sciences

Department of Computer Science

Dean of Faculty

Associate Professor, PhD Zhamanov A.

« 30 » May 2022

Topic of the thesis:

Weather-induced flight delay prediction using Machine Learning

Thesis submitted as part of the requirements for the award of the MSc in "7M06102 - Computer Science", SDU, 2020-2022

Head of Department	PhD, Cemil Turan
Academic Supervisor	PhD, Andrey Bogdanchikov
Master's student	Yerkebulan Zhumagali

Kaskelen, 2022

Abstract

Many businesses depend on different airlines to link them to other parts of the world, and the aviation industry today plays a crucial role in the global transportation sector. Extreme weather, on the other hand, might cause flight delays, which can have a direct influence on airline services. To solve this issue, accurately projecting flight delays enables passengers to be well prepared for their journey's interruption and allows airlines to react to anticipated causes of flight delays ahead of time, decreasing the variety of consequences. as a consequence, airlines and experts are focusing their efforts on cutting down on flight delays. It is vital to predict flight delays in order to reduce aircraft delays. Establishing a reliable and accurate flight delay prediction system may provide decision-makers with a clear path to make effective scheduling choices. Weather is a most common reasons of aircraft delays, and subsequently it linked to other categories. The NaS type, for example, might include delays caused by fly rerouting due to bad weather. Late-arriving aircraft is also influenced by weather, but airlines do not define the reasons as weahter. When these factors are taken into consideration, weather over 40% of late minutes are accounted for by this factor. as a result, assessing the impact of bad weather on carrier delays is critical for smooth flight operations.

Андатпа

Көптеген кәсіпорындар оларды әлемнің басқа бөліктерімен байланыстыратын әртүрлі авиакомпанияларға тәуелді және авиация саласы бүгінде жаһандық көлік секторында шешуші рөл атқарады. Екінші жағынан, төтенше ауа-райы рейстердің кешігуіне әкелуі мүмкін, бұл авиакомпанияларға қызмет көрсетуге тікелей әсер етуі мүмкін. Бұл мәселені шешу үшін рейстердің кешігуін дәл болжау жолаушыларға сапарларының үзілуіне жақсы дайын болуға мүмкіндік береді және авиакомпанияларға рейстердің кешігуінің күтілетін себептеріне алдын-ала жауап беруге, салдардың алуан түрін азайтуға мүмкіндік береді. Нәтижесінде авиакомпаниялар мен сарапшылар рейстердің кешігуін азайтуға күш салуда. Ұшақтардың кешігуін азайту үшін рейстердің кешігуін болжау өте маңызды. Рейстердің кешігуін болжаудың сенімді және нақты жүйесін құру шешім қабылдаушыларға кестені тиімді таңдауға нақты жол бере алады. Ауа-райы ұшақтардың кешігуінің ең көп таралған себептерінің бірі болып табылады және ол басқа санаттармен де байланысты. Мысалы, ұлттық ұшу шеңберлерінің санатына ауа-райының қолайсыздығына байланысты ұшу маршрутының өзгеруіне байланысты кідірістер кіруі мүмкін. Ауа-райы ұшақтардың кешігуіне әсер ететін фактор болып табылады, дегенмен авиакомпаниялар ауа-райының себептерін атамайды. Егер осы факторларды ескеретін болсақ, ауа-райына кешіктіру минуттарының жалпы санының 40% - дан астамы келеді. Нәтижесінде, қолайсыз ауа-райының тасымалдаушының кешігуіне әсерін бағалау рейстерді үздіксіз орындау үшін өте маңызды.

Аннотация

Многие предприятия зависят от различных авиакомпаний, которые связывают их с другими частями мира, и авиационная промышленность сегодня играет решающую роль в глобальном транспортном секторе. С другой стороны, экстремальные погодные условия могут привести к задержкам рейсов, что может оказать прямое влияние на обслуживание авиакомпаний. Чтобы решить эту проблему, точное прогнозирование задержек рейсов позволяет пассажирам быть хорошо подготовленными к прерыванию их путешествия и позволяет авиакомпаниям заблаговременно реагировать на ожидаемые причины задержек рейсов, уменьшая разнообразие последствий. Как следствие, авиакомпании и эксперты сосредоточивают свои усилия на сокращении задержек рейсов. Крайне важно прогнозировать задержки рейсов, чтобы сократить задержки самолетов. Создание надежной и точной системы прогнозирования задержек рейсов может предоставить лицам, принимающим решения, четкий путь к эффективному выбору расписания. Погода является одной из наиболее распространенных причин задержек самолетов, и она также связана с другими категориями. Например, в категорию Национальных рамок полетов могут входить задержки, вызванные изменением маршрута полета из-за плохой погоды. Погода также является фактором, влияющим на опоздание самолетов, хотя авиакомпании не называют причины погодой. Если принять во внимание эти факторы, на погоду приходится более 40% от общего количества минут задержки. В результате оценка влияния плохой погоды на задержки перевозчика имеет решающее значение для бесперебойного выполнения рейсов.

Contents

Nomenclature

ATFM Air Traffic Flow Management

CAAC China Civil Aviation Administration

LAAD Late arriving aircraft delay

TFM Traffic Flow Management

WITI Weather-Impacted Traffic Index

1. Introduction

1.1 Motivation

In recent years, several studies have focused on flight delay. As the demand for air travel has expanded, flight delays have increased. Approximately 20% of all planned commercial flights are delayed, in accordance with Bureau of Transportation Statistics [33]. Airline delays cost airlines a lot of money every year and give customers a great deal of inconvenience. Airline delays are classified as follows by BTS:

- Air carrier;
- Extreme weather;
- National Aviation System;
- Late-arriving aircraft;
- Security.

Weather is not just one of the most common reasons for delays, but it also affects other categories. Airline delays due by adverse weather rerouting, for example, may fall under the National Aviation System category. When all of these factors are taken into account, weather over 40% of late minutes are accounted for by this factor [16].

Individual flight delays were forecasted using supervised machine learning algorithms that took into account factors flight schedules and weather forecasts for departure and arrival are also included. Improving forecast capabilities, Rather than the complete National Airspace System, Individual departure and arrival models were created by gathering meteorological data originating from geographic location at each airport.

Weather has a major influence on airport operations as well as the overall functioning of the aviation network. Due to extreme weather conditions, airport capacity limits may force operations to be delayed. To ensure timely operations, it is necessary to forecast aircraft activities over their whole paths. Uncertainties during the flight's airborne period have a modest influence on total timeliness.

As a result, studying the effects of adverse weather on air operations delay is essential in order to perform successful operations in the air. Furthermore, a decision-making aid based on the research might alert customer and companies ahead of time of weather-related delays, thus decreasing financial losses. In this study, supervised machine learning algorithms were used to evaluate individual airline arrival delays. Machine learning can be used for a variety of reasons. To begin with, there is just too much historical flight and meteorological data to be evaluated analytically. Furthermore, interactions between causal factors and delays, as well as cross-factor correlations, are exceedingly complicated and nonlinear, making it impossible to evaluate all hypotheses.

Data analytics is now being pushed forward by the ongoing rise in storage capacity and processing power. Companies (particularly those with a high IT component) are gathering vast amounts of data (also known as Big Data), including online logs, client information, manufacturing and sales monitoring, and so forth.

Data mining methods, for example, may be used to extract information from these datasets that might assist a firm obtain knowledge (for example, about consumer behavior) or utilize the information as a foundation for new goods or services.

To do this, machine learning methods might be utilized. The study of algorithms that try to learn from data and generate predictions is known as machine learning (ML). These algorithms may be utilized to create the appropriate predictions as long as flying data is available. ML has applications in a variety of domains, including economics, quantum physics, and medical imaging. There are three types of learning algorithms: supervised, unsupervised, and reinforcement learning algorithms. The purpose of supervised learning is to develop a function that maps input variables to output variables based on input-output combinations that have been trained. If this function must map the input to one of a small number of categories, it is a classification issue. It's a regression issue if the output is a continuous variable. In contrast to supervised 4 learning, goal values are not specified in an unsupervised learning situation. Without knowing the output variable, the objective is to learn a function that describes the structure of the input data. Finally, reinforcement learning is concerned with how systems should respond in order to maximize a certain form of reward [7]. Because the available input-output pairings may be used to train a model, and we're looking for a function that can translate input variables to a continuous output variable, predicting delays is an unsupervised ML regression issue. The emphasis of this thesis is on forecasting domestic airline arrival delays in the United States in 2015. As a result, the research question is: Which machine learning system can most accurately anticipate domestic flight delays in the United States in 2015. A data collection from BTS including observations of domestic flights in the United States in 2015 was utilized to address this research issue. We use flight data, meteorological data, and aircraft characteristics to create a set of factors that might explain the arrival delay. Models are trained using three distinct ML strategies after pre-processing the data (e.g., dealing with missing values). Multiple linear regression is the first approach employed, with two distinct techniques of reducing the associated error function (note that this is actually an econometric technique). The second is a support vector regression-based machine learning method (SVR) that employs a smooth version of the error function often employed in SVR. Neural network regression is the third and final ML approach. The relevant hyperparameters are tuned for each model. Finally, model performance is assessed by looking at how effectively the model generalizes to new, previously unknown data. We examine the model results to see which machine learning technique is the best accurate at predicting arrival delays.

Improving the performance of airports and, more broadly, the National Airspace System (NAS) has long been a focus of the commercial aviation sector, as well as Flight delay modelers and analysts are interested in this issue. Over 44,000 commercial flights operate on a daily basis in the United States along with the capacity constraints that must be met to allow these flights to take place, create a dynamic and complicated system for air transport. As a result of this, it is far more difficult for air traffic controllers and airline employees to do their duties. to deal with unanticipated traffic generated by adverse weather and react to demand in a proactive manner.

For air traffic control, airlines, and customers, flight delays may have major economic consequences. This is because delay is a key performance indicator for any system and is costly to airport stakeholders. The severity of flight delays and the related costs The earliest indications of an imbalance in the airspace system are for ailrines and passendasadsgers. The continuing expansion in passenger traffic and the existing demand-capacity mismatch are causing an increase in aircraft delays airpdasdasorts and routasde secasdastors, especially those alreadasdady operating at capacidasty, during peak hours. Improved decision-making tools, such as more systdasematic delay monitoring utilizing very exact prediction models, are crucial for ensuring optimal trafdasdfic flowdasds.

When it comes to making decisions in air trafdasfic contdasrol (ATC), it is necessary to weigh several different aspects that are specific to each situation. An extensive amount of data processing is necessary in order to arrive at a comprehensive, strategic, and numerical judgment. It is possible for dispatchers and controllers to make more effective tactical and strategic decisions in order to reduce delays and improve the operation of the NAS. A multi-processor decision-making technology that allows all stakeholders in the aviation industry to exploit big data is the future of NAS develodaspment. Developing models that can effectively forecast latency and analyze external variables that have a significant impact on NAS performance is thus of interest and highly sought after in this situation.

Using supporsdat vecdator machindsaes, you may foresee and evaluate important elements that affect when an airplane will arrive. delays in this research (SVM). Statistical logistic regression analysis was used in this instance to give early insight into possibly statistically relevant components as well as a comparison of the model's predicted performance. Studies that employ Analysis of aircraft delays is often overestimated by machine learning algorithms. to alter hyperparameters to enhance forecasting performance when it comes to forecasting. As a consequence, it looks at the usage of a metaheuristic technique in hyperparametric tuning of an SVM model for latency prediction, as well as delving further into the model findings. Furthermore, machine learning research often concentrate on prediction rather than variable effects. For the first time, both of these components are discussed in this essay.

As a consequence, this research aims to identify and explore the structure of factors that influence the arrival delays of airplanes. There has never before been an attempt to use ML models with discrete choices and supervised learning to explore delays in airplane arrivals. So the study's main contributions include: • examining the viability of using SVM model for flight delay analysis; and • evaluating the effect of contributing components.

• Look at metaheuristic optimization's potential to increase SVM prediction performance by tweaking parameters.

Ufassing logistdsasic regressifason and macadshine learafsning, an in-depth analysis of the factors that influence the arrival time of a flight was undertaken.

Even though this research approach was first developed to investigate airline arrival delays, it may be used to various areas of air transportation research for data-driven analysis.

1.2 Aims and Objectives

Airlines have evolved from basic contract letter carriers to cognitively compelling corporations during the last century. However, the airline sector is highly competitive, dynamic, and unpredictable, which creates uncertainty. One such hazard is right delay, which can be caused by a variety of circumstances including severe weather, physical laws, delayed right arrivals, and crew-related concerns [36] . It is self-evident that flight delays have a direct influence on passenger pleasure and, as a result, airline profitability. As stated by the sum of delay affect, the total cost of all air transportation delays in the United States in 2007 was 32.9 billion [6]. As a result of this gure, airline businesses have been preoccupied with comprehending and limiting delays. Many academics have tried to foresee these uncertainties in order to assure more trustworthy and robust results. To achieve these objectives, new approaches to airline planning must be developed, which may be done by taking operational difficulties into account during the planning phase. Flight planning is a complex procedure with several factors. Scheduling, as one of the duties in right planning, is done in a step-by-step procedure that includes two phases, namely the planning and operational phases. Certain low-complexity sub-problems appear in each step and are handled consecutively [8]. There are four phases in the planning phase: right scheduling, feet assignment, feet routing, and crew assignment. Each of these processes is completed independently and then fed into the next. Revenue management, gate assignment, and irregular operating processes are all part of the operational phase. One of the most significant aspects of the operational phase is revenue management, which is utilized by airlines to limit the number of seats sold at various rates. The allocation of gates to boarding and departure is the subject of gate assignment. The control tower assigns gates to an airplane as it lowers to a height where it is cleared for landing based on the activity of other flights. The last phase is irregular operation, which refers to how often light program implementation deviates from planned. When one right is delayed, the airline operation control center takes steps to prevent the delay from spreading to other rights. Three ways are used to assure an effective right planning process: the traditional approach, the robust approach, and disruption management. The difficulties that arise in each step of the planning phase are executed independently in the traditional method, or the problems in a number of phases are merged and treated concurrently in the classic approach. [7] [30] provides more information on this method. The difficulties that arise in one or more phases of the planning phase are treated in connection to the operational phase in the resilient method [26] [2] provides more details on this strategy. The consideration of operational difficulties that arise during right planning is known as disruption management. This strategy is employed when a disruption occurs, as the name implies [25] [17].

1.3 Thesis Outline

An introduction, five chapters, including a conclusion, a list of references, and an abstract in three languages comprise the dissertation work. Introduction Introduction is the first chapter. The significance and relevance of the dissertation subject are established in this chapter, as are the aims of the dissertation work and the tasks to be completed, as well as the work's scientific innovation and practical value. A chapter-by-chapter overview of the work is provided. Chapter 2 examines the topic area and existing methods for proper prediction, determining their benefits and drawbacks in terms of flight delay prediction. It also discloses the necessity to rectify the allowable prediction mistakes, gives categorization kinds and analyses of available correction procedures, and explains the needs for the system under development. The methodologies employed and the created approach for automatic correction of identification mistakes based on normalization are described in Chapter 2. In addition to related work with the anticipation model and show relevance method developed in other models.. The methodology and architecture are described in Chapter 3. It depicts the steps of flight delay prediction installation, specifies the system's order, and provides the method for discovering and prediction flight delay. This chapter contains information about different algorithms, which has been explored for prediction model analysis. The data and datasets used to test the technique are presented in Chapter 4. The fifth chapter outlines the circumstances and process for testing the developed system, as well as the criteria for evaluating prediction quality. The work results for each algorithm are displayed, and it is determined which method is suited and has benefits over others. The results of an experimental evaluation of the suggested prediction approach are provided, with an example discussion. In the conclusion, the work's findings are summarized, as well as the research's primary findings, difficulties with solutions, and future directions for scientific research.

2. Literature review

Researchers and analysts are primarily concerned in predicting the causes of flight delays, hence they have focused their efforts on gathering data on flight and weather.

2.1 Basic representation of flight delay prediction

2.1.1 Flight delay definition

The term "flight delay" refers to the situation in which the carrier is unable to depart from or return to the station in accordance with the terms of the transportation contract when transporting passengers by air. The flight schedule or the time mentioned the on be should ticket air used as direct proof of flight delay. It may be classified into two categories based on the cause for the delay: fair and unreasonable. In general, the former will not be responsible for compensation; the latter will be responsible.

A flight delay is described as a flight departing or arriving later than planned, which is common in most airlines throughout the globe, resulting in significant financial losses for the airline and significant discomfort for passengers.

According to the China Civil Aviation Administration (CAAC), extreme weather accounts for 47.46% of delays, while air route issues account for 21.14%. Air traffic control and other factors account for 2.31% and 29.09%, respectively, due to airline business difficulties or technological issues. Recent research has focused on developing a reliable method for predicting the likelihood of a flight delay or the length of the delay in order to better use ATFM [3] to decrease the amount of delay. Modeling the prediction model may be done in two ways: classification and regression.

Many recent research have used machine learning approaches to improve classification models and have shown encouraging results For example, L. Hao et al. utilized a regression model to estimate flight delays at New York's three main commercial airports. However, present approaches are limited in their ability to improve the accuracy of flight delay prediction for a variety of reasons. The following are the reasons for the flight delay: the variety of causes, the complexity of the causes, the relevance of the causes, and the inadequacy of accessible flight data.

A flight delay is defined as a circumstance in which the carrier is unable to depart from or return to the throughout air transportation, the agreed-upon station. The two forms of flight delays are departure and arrival delays. There are three types of delays: those in which the departure time is delayed but the final arrival time is not; those in which both times are delayed; and those in which both times are delayed. The term "delay" is used to describe any of these situations.

The inability of the airline to depart the station on time as specified in the schedule or on the ticket is referred to as a "departure delay." The word "arrival delay" refers to a carrier's inability to arrive at the station on time as stipulated on the schedule or on the air ticket. There is no provision in the current legislation for establishing the arrival time. Some academics believe that the departure time should be calculated from the moment the plane's door is unlocked. According to the author, he or she agrees with this.

In the airline sector, delays have been proven to build as the day goes on. For example, in the summer, airlines are able to maintain their flights on schedule for roughly 85% of the time, and sometimes even better until mid-morning. However, by mid-afternoon, this rate had declined to a low of 70%, and by evening, it had depreciated to around 60%.

According to Dieterich Lawson et al, flights are delayed or canceled due to air traffic control assistance, technical difficulties, and weather. Consideration of multiple classifiers for flight delay prediction. Dieterich Lawson and William Castillo used SVM to undertake exploratory research on single trip datasets from San Francisco International Airport for up to a year, allowing them to use unique airport attributes rather than a cluster. They couldn't come up with a convincing rationale to keep going, so they tried something different. The authors discovered that SVM did not produce the expected results, namely a 14.4% minimum error rate. Given the mistake rate, expected accuracy and recall were definitely low, especially given all flights were supposed to arrive on time. Because the SVM classifier's accuracy was inconsistent, the authors chose to employ a non-SVM classifier instead. They then employed a Naive Bayes approach on top of Apache Hadoop. Naive Bayes, which is rather fast, was the only classifier they were able to train on the whole 135 million row dataset.

Airline operations are extremely complicated procedures aimed at regulating a large number of expensive, closely limited, and interdependent resources, such as crews, planes, airports, and repair facilities. [23] Although many research have been conducted on airline planning issues, few have been conducted on the characteristics of airline delays and the prediction of delay data. When anything happens later than it was planned, scheduled, or expected to happen, it is called a delay. Bad weather, seasonal and holiday demands, airline policies, technical concerns such as problems with airport infrastructure, luggage handling, and mechanical gear, and the accumulation of delays from previous flights can all cause delays in departure. (Figure 2.1)

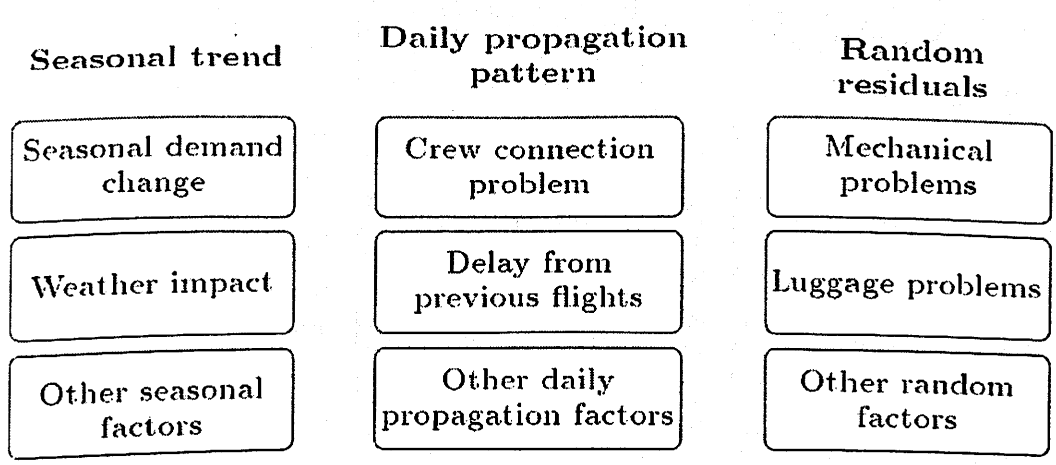


Figure 2.1: Aspects impacting to departure delays

2.1.2 Early attempts to calculate flight delay probabilities

Transportation via air is a highly complex industrial process. Flight delays may be extremely problematic due to the high-tech and risky nature of aviation transportation. According to Civil Aviation's Standard Statistics Method, flight delays can be caused by a variety of things such as in-flight issues like mechanical failures or weahter delays as well as other factors like flow conrtol and maintenance engineering as well as airport infrastructure and other aspects like flight guarantee and crew. They're all just facts that don't differentiate between carrier and non-carrier causes, subjective and objective factors. It simply means that the explanations are stated without any blame being assigned.

Reasonable delays are often caused by external circumstances that cannot be controlled, such as the weaegher, the manaegement of air traffic, flow cotrrol, accidents, and secieity checks. Flight allocation, maintenance, material assurance, and flight crew delays, on the other hand, are deemed unjustified delays since they should have been anticipated and prevented by the airline.

The words "reasonable delay" and "fair delay" are interchangeable. The existing state of China's air transporerdastation must be taken into account. A reasonable delay time should be supplied as a transition phase to balance the benefits of air carriers, and the time period should be adequately decreased until it is canceled. Universal reasonableness should be the criterion for a suitable delay time definition.

The airline's advertised flight departure and arrival timings are the most important reference points. Second, the passenger's reasonable expectation of airline convenience should be considered. The majority of domestic flights in China may be completed in less than two hours. It will be absurd to wait more than two hours. Regardless of the cause of the delay, airlines cannot be exempt from the law's notice and restitution responsibilities.

Flight delays are mostly economic in nature and effect passengers, airlines, and airports. Due to the unpredictability of their occurrence, travelers frequently seek to arrive several hours early for their appointments, increasing their travel expenditures. Airlines face penalties, fines, and increased operating costs such as employee and aircraft delay at airports. Delays may have an environmental effect by increasing fuel use and gas emissions.

International aviation transportation conventions make broad statements about flight delay and responsibilities to be assumed, from the convetnion Warwas, but they do not define the term "flight delay." When determining the cause of a flight delay, the most important factor to consider is whether the carrier completes the transportation within the contract's time frame. It typically refers to the time specified in the transportation contract, as well as the time shown on the carrier's itinerary or on an airline ticket. Because the word "regular flight" is defined, it simply means a flight having a set departure time.

When a carrier fails to convey people or products to their destination on time, a flight delay occurs.

The recognition of air flight delays, their causes and association with bad weather has been the focus of researchers for many years, especially since the late 1990s, with organizations such as MITER, MIT Lincoln Lab, NASA, and academic institutions such as MIT. , University of Maryland and George Mason University. While the original terminal weather considerations were expanded into the year-round NAS operating model, the focus was on the effects of convective weather. Airport delay models were trained using historical traffic and weather data as well as various regression approaches and then tested against actual delay data in a quasi-prediction mode. It was agreed that the development of a more complex model was worthwhile due to the prevailing focus on convective weather effects and the generally poor detail of the impact of non-convective weather components on airport operations. In addition, the new generation of convective weather forecasting products - which our software tools and models can accept and convert into TFM constraints - opens up new possibilities for more accurate delay prediction. [24] WITI is based on current weather conditions and expected traffic demand. WITI-FA, its twin, is based on forecast weather and traffic demand. WITI is a weighted average of three factors. The en route component (E-WITI) takes into account the impact of convective weather on the airport's main routes. The Terminal Component (T-WITI) measures the capacity loss due to surface weather, which is proportional to the number of airport operations. The queuing delay (DELAY) component calculates the overall impact of excess bandwidth demand on traffic. WITI-FA contains similar components that depend on weather forecasts.(see Figure 2.2)

$$\boxed{\textbf{Overall WITI} = \text{EWITI} * \mathit{Wght\_Coef\_EWITI} + \text{TWITI} * \mathit{Wght\_Coef\_TWITI} + \text{QDelay} * \mathit{Wght\_Coef\_Qdelay}}$$

Figure 2.2: The Three-Component WITI

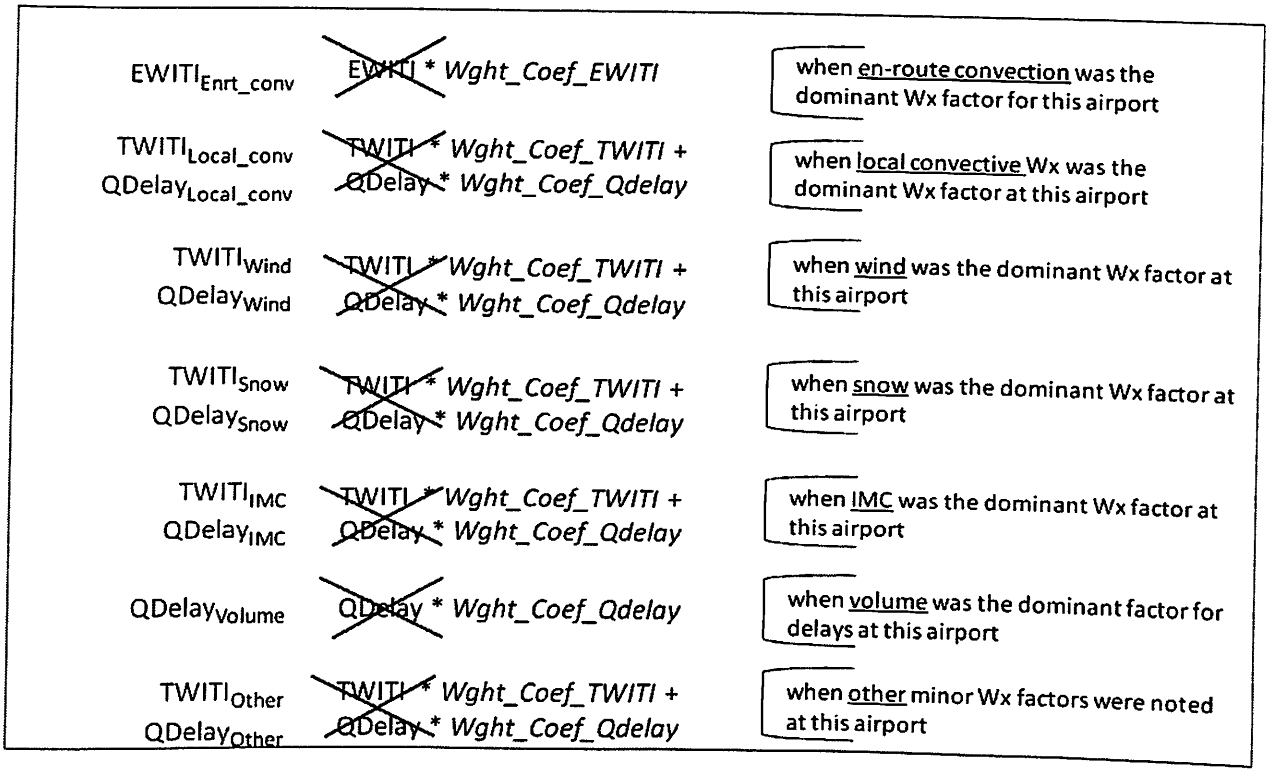


Figure 2.3: The 12-Component NAS WITI

The above-mentioned seven weather/traffic elements have now been transformed into a 12-component Airport WITI. The components include an En-route WITI (EWITI), which is unaffected by terminal weather, and a Volume WITI, which is unaffected by weather and just depends on traffic. Local convective, Wind, Snow, IMC, and Other are translated into two WITI components: T-WITI (linear) and Q-DELAY (nonlinear).

The original three-component WITI is shown in Figure 2.2. Figure 2.3 displays the 12-component Airport WITI with weather factor specific weights, where we are now talking about a "Snow T-WITI," "IMC T-WITI," or "Wind T-WITI" instead of a generic T-WITI. Finally, airport-specific meteorological variables and weights are used to transform "Snow TWITI" or "IMC T-WITI" into "JFK Snow TWITI" or "ORD IMC T-WITI" or "DFW Wind TWITI." That is, the 12 components and their weighting coefficients are now linked to a specific airport, as well as its individual weather conditions and how they affect it (2.4).

Due to building a predictive model and the need to obtain the last full year of airport activity, we typically use current data to calibrate the model in this experiment, as it is designed to predict delays. In addition, due to the operation of runways at several airports, previous data was found to be inadequate. Based

Every hour $[i]$, for airport $[k]$:

$$
\begin{aligned}
\text{Airport}_k\ \text{WITI}^i = {} & \text{Airport}_k\text{\_EWITI}^i * \textit{Wght\_Coef\_EWITI}_k\ + \\
& \text{Airport}_k\text{\_TWITI}^i{}_{\text{Local\_conv}} * \textit{Wght\_Coef\_TWITI\_Local\_conv}_k\ + \\
& \text{Airport}_k\text{\_QDelay}^i{}_{\text{Local\_conv}} * \textit{Wght\_Coef\_Qdelay\_Local\_conv}_k\ + \\
& \text{Airport}_k\text{\_TWITI}^i{}_{\text{Wind}} * \textit{Wght\_Coef\_TWITI\_Wind}_k\ + \\
& \text{Airport}_k\text{\_QDelay}^i{}_{\text{Wind}} * \textit{Wght\_Coef\_Qdelay\_Wind}_k\ + \\
& \text{Airport}_k\text{\_TWITI}^i{}_{\text{Snow}} * \textit{Wght\_Coef\_TWITI\_Snow}_k\ + \\
& \text{Airport}_k\text{\_QDelay}^i{}_{\text{Snow}} * \textit{Wght\_Coef\_Qdelay\_Snow}_k\ + \\
& \text{Airport}_k\text{\_TWITI}^i{}_{\text{IMC}} * \textit{Wght\_Coef\_TWITI\_IMC}_k\ + \\
& \text{Airport}_k\text{\_QDelay}^i{}_{\text{IMC}} * \textit{Wght\_Coef\_Qdelay\_IMC}_k\ + \\
& \text{Airport}_k\text{\_QDelay}^i{}_{\text{Volume}} * \textit{Wght\_Coef\_Qdelay\_Volume}_k\ + \\
& \text{Airport}_k\text{\_TWITI}^i{}_{\text{Other}} * \textit{Wght\_Coef\_TWITI\_Other}_k\ + \\
& \text{Airport}_k\text{\_QDelay}^i{}_{\text{Other}} * \textit{Wght\_Coef\_Qdelay\_Other}_k
\end{aligned}
$$

Figure 2.4: Airport Specific 12-Component WITI

on the received data, we determine the preferred runway configuration priority for each airport. For the current weather conditions at the airport, the model selects the first available runway configuration. Capacity for each configuration is calculated using historical arrival and departure statistics. The capacity of each runway configuration is calculated using values close to the right side of the distribution of actual arrivals and departures. Where FAA bandwidth standards are available, these bandwidth reduction predictions are compared with them. As a measure of throughput, we effectively use real throughput under near-ideal conditions with heavy scheduled traffic. Then, using historical data, power losses are calculated under certain weather conditions. Airport capacity for arrivals and departures is reduced by a percentage that reflects deterioration caused by prevailing weather conditions (eg thunderstorms, high winds, etc.). The model takes into account differences in operating procedures at different airports using airport-specific capacity reductions.

Fitting a multivariate linear regression model with each component of WITI as the independent variable in the model determines the relationship between WITI and latency. The dependent variable is the total delay (in minutes). The sum of arrival and departure delays is the total delay; Arrival delays are calculated by multiplying the number of arrivals at each airport by the average airport arrival delay per hour, and departure delays are calculated in the same way. The weights assigned to each component of the WITI model are the coefficients for each explanatory variable in the bestfit model.

2.2 Overview of actual existing methods

2.2.1 Mixed approach

The a multi-tagged classification technique that includes modules for delayed departure and arrival prediction, is first given a feature selection process. Weather data and air traffic performance statistics are two known factors that might affect air traffic. We propose an algorithm based on a combination of Bureau of Transportation Statistics, National oceanic and atmospheric administration, and aviation System Performance Metrics to determine the ideal training features that accurately forecast departure and arrival delays. Second, a delay propagation model is developed, with one important parameter, which affects to occurring Late arriving aircraft Delay (LaaD). The LaAD describes the inter-aircraft relationship between a previous arrival and a current departure. Last but not least, the LaaD is used to create a chain model that connects earlier arrival delays to current depatriation delays. Measures arrival delay, computes LaAD, and forecasts future departure delays together with the original departure delay. The model's accuracy might be improved even more by updating the actual detention delay with the sequence number, as described above. [15] In addition to the pre-departure and post-departure delayed departures, the pre-arrival, pre-departure, and post-departure delays are all part of the mixed approach for forecaching individual aircraft flights. Figure 2.5 depicts the hybrid method's structure. It is linked to the departure delay predicitions through the departure predicitions model. The chain model may be created by repeatedly running the associated modules once the initial delay has been supplied. The flight schedule and training set components are the last two inputs. The arrival and departure delay prediction modules are based on a random first model that has been trained with certain properties. If you've worked with previous LLAADs, you'll notice that the delay propagation module has been fine-tuned to perform properly.

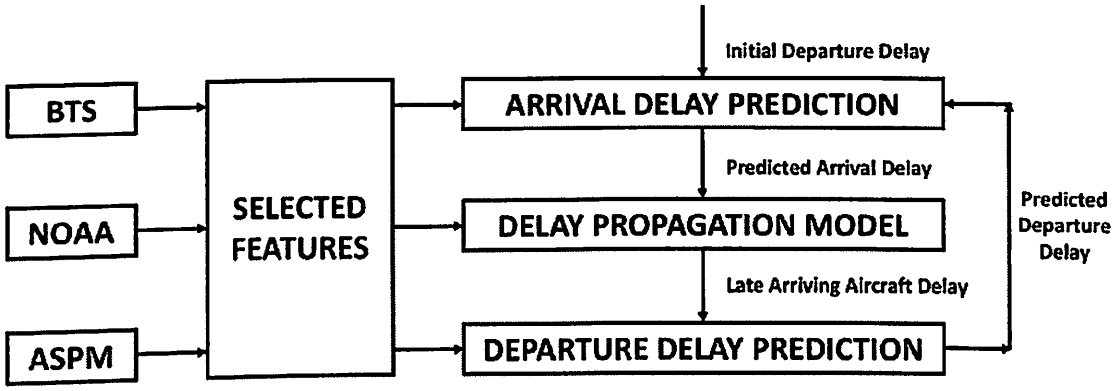


Figure 2.5: A hybrid strategy for linked delay prediction

The Random Forest is a multi-decision tree ensemble approach [12]. By combining the Bootstrap aggregating algorithm with RF, the shortcomings of individual decision trees are addressed. [21] and random space approach [22]. Because it has minimal sensitivity to outliers in training data and can detect high-dimensional data quickly, it performs well. RF is widely used in industry [21]. Two considerations led to the decision to make radio frequency (RF) the core of our prediction modules. To begin with, it has been shown that RF outperforms other categorization approaches [29] [16]. Second, in its learning process, RF may output the relevance of the characteristics.

There are two ways that delays may spread across the air traffic system: first, they can spread via the en route link that connects airports, and second, they can spread on the ground inside the airport itself. Although the primary RF classifier is able to predict individual delays in departure and arrival, the transition from the actual arrival delay to the LaaD is not an easy one. According to BTS, the LaaD is one of the five primary causes of airline delays. It describes the intrinsic link between the prior arrival delay and the current departure delay on the same aircraft. The LaaD is one of the major reasons for airline delays.

The CH to Chi definition was created by researchers to test the impact of the flight iteration number in the chained delay model, where i is the number of preceding real-world depatriation departures that have been updated to work with the chained delay model. i has a range of one to 10 iterations.You are able to see the impact of the iteration number on the simulation result by adjusting the iteration number of the flight data that is fed into the model. When looking at the model associated with the teration number, two criteria—relaxed accuracy and accuracy—are specified. The diagram depicts the distribution of relaxed accuracy and accuracy as a function of teraton number. Also shown is the relaxed accuracy as a function of accuracy. The model's precision, as well as its lax accuracy, are both impressive., increases with each iteration as the true departure delay is included into it. Due to the fact that the total number of data points does not change, the number of data points that fall within the relaxed accuracy zone (error groups 0 and 1) will increase, while the number of data points that fall within error groups 2 to 10 will decrease. When the number of rounds increases, the prediction error also increases, which causes the overall number to decrease at a faster rate. Convergence speeds decrease as the number of iterations increases. all of the criteria converge as the number of iterations grows. To get a clearer picture of the erroneous dstribution of a predicate, divide 16 by the number of predicate errors and divide by the number of predicate errors to get at the predicate error dstribution. In order to make the high error classification distribution more straightforward to see, the number of erroneous accounts in the high error classification is much too little to be noticed when compared to those in the lower error classification. As the number of iterations rises, the error class of 0 and 1 — also known as the relaxed accuracy zone — will expand somewhat. The other class, on the other hand, will shrink as the number of iterations increases. As the mistake class becomes larger, the precision error decreases. This also shows that changing the actual deposition time improves the accuracy of the forecast and reduces the prediction error.

2.2.2 Machine learning algorithms with SMOTE technique

When the categorization categories are not roughly equally represented, the dataset is described as unbalanced. [14] The training data is unequal based on this criterion, since there are three to four times as many on-time flights as there are delayed flights. It means that even if every flight is labeled as on-time, a forecast accuracy of more than 75% is possible. Identifying the minority class is crucial in the real world because misclassifying instances from the minority class is expensive. [20] To solve concerns arising from skewed data, To change the training data is unequal according to this criterion since there are three to four times as many on-time flights as there are delayed planes instances should be generated.

SMOTE is a method for making fake cases of minority classes through over-sampling. By injecting synthetic cases along the line segments connecting any/all of the k minority class nearest neighbors, the minority class is over-sampled. SMOTE is an effective solution for unbalanced datasets that addresses over-sampling and under-sampling issues. SMOTE has been found to boost classifier performance in a variety of application domains [11] [13]. In the training step, because it yielded more accurate estimates than other sampling strategies, the SMOTE method combined with random under-sampling was selected as the method of choice.

Figure 2.6 provides a high-level picture of the model used to predict individual flight delays. The model's two most important components are the training and prediction phases. The gathering of data is the first step in the process of training an employee. As the join keys, the scheduled departure time as well as the airport are employed in order to connect historical flight data and weather data. Both the estimate of missing data and the normalization of the data are done during the preprocessing step of the process. Training sets are then completed and utilized to train the prediction model through sampling procedures. The same data and preprocessing steps are used for the prediction set as they were for the training set. It's then put into a model that's been trained on the data. At this phase, the model assigns a label to each individual data point. The methodologies that were used are broken down in great detail in the preceding section, and the following section will go on to explain the remaining aspects of the model.

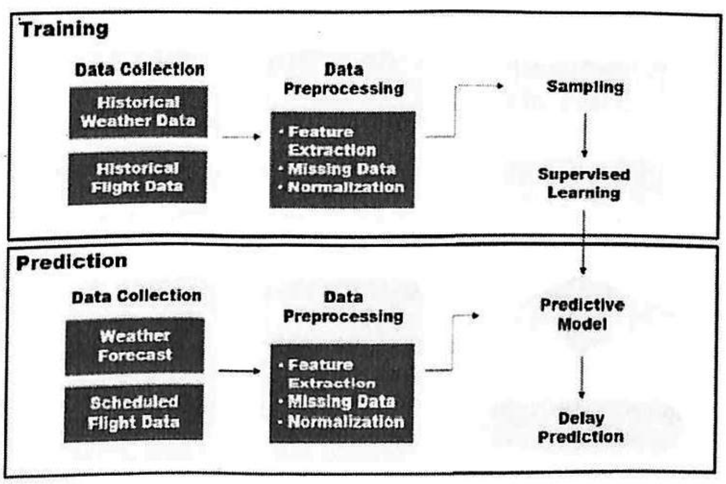
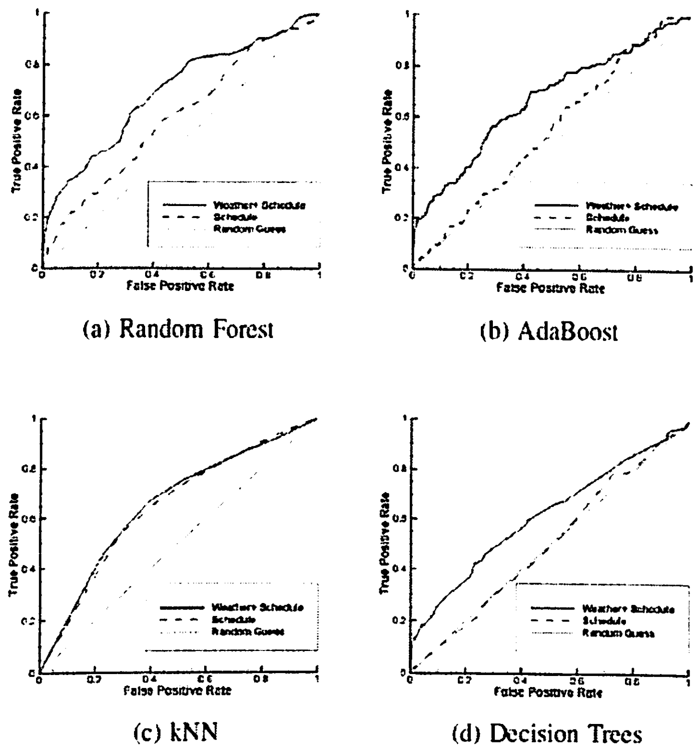


Figure 2.6: Summary of the model



(a) Random Forest (b) AdaBoost

(c) kNN (d) Decision Trees

Figure 2.7: ROC

First, the effects of meteorological data on four models' prediction ability are explored. A red dotted line in Fig. 2.7 A model that was trained using both weather and schedule data is represented by a solid line that is blue in color, and a model that was trained using both weather and schedule data is reflected by a solid line that is red in color. Because a diagonal line represents random guessing, the closer the ROC curve is to the diagonal, the less accurate it is. Weather data boosted each model's ability to forecast the future since blue solid lines are closer to the optimal spot on the ROC curve than red dotted lines. AdaBoost and decision trees performed nearly as well as a 'random guess' without meteorological data in terms of prediction accuracy. Without meteorological data, AdaBoost and decision trees are unable to predict whether or not a scheduled aircraft will be delayed. Compared to other classifiers, kNN has a very small gap between two curves with and without weather data. Dimensionality's curse is to blame. Many possible combinations of components are available in the flight data. There are 56 unique characteristics in the flight data. In such high-dimensional data, measuring distance between sample locations becomes useless [28], It is not useful for kNN to boost prediction power by integrating weather data with 80 features.

It will be easier to account for local weather occurrences in network-wide as-

Classifier	Accuracy (%)	Time
Random Forest	81.37	8
AdaBoost	78.05	12
kNN	61.69	2
Decision Trees	77.02	0

Table 2.1: Accuracy of training and time

sessments when weather influences on airport operations are categorized. Weather data from meteorological reports is correlated with airport performance data such as aircraft plans with planned, actual, and delayed movements using machine learning algorithms. Using recurrent and convolutional neural networks, unsupervised learning was used to cluster performance affects at the airport and identify the corresponding meteorological data. As a result, our machine learning system is able to match decreased airport performance during local weather events correctly as well as anticipate delays based on weather forecasts and flight schedules, respectively. As a starting point for future research, this study shows the efficacy of machine learning algorithms for categorisation. We may get a better grasp of air transportation's local and nedasstwork-wide intsaderdependencies by modernizing the present expert-based weather classifications.

To compensate for the anticipated aircraft delays, air traffic control may alter flight plans by, for example, trading slots, establishing a new runway, or modifying the runway design. External occurrences, such as runway and airspace restrictions, might make a flight plan outdated, which must be taken into consideration while attempting to minimise time. Messages, on the other hand, signal changes in environmental circumstances and alert users to external occurrences. METARs provide fundamental meteorological data such as air temperature as well as information on unusual weather events such as thunderstorms and hail. NOTAMs, for example, describe runway or airspace closures. In addition, generic airport and runway information, such as the number of runways or the length of a runway, might help with delay planning. Changes in the environment and external events such as runway and airspace closures must be recognized as soon as feasible by air traffic control in order to alter flight plans and minimize delays. Air traffic control systems now in use do not take full use of the vast amount of data available for detecting impending congestion and, as a result, aircraft delays. As a result, in air traffic situations with a significant risk of delay, flight plans are not adjusted quickly enough. In this context, the word "air traffic scenario" refers to an ensemble of flights leaving or arriving at a certain airport during a specific time period, as well as information on the aircraft involved (number of small, medium, and large aircraft).

To investigate the impacts of sampling procedures, the model is trained with and without sampling approaches, and the prediction performance of the two situations is compared. Tables show the accuracy and elapsed time of four classifiers that were trained using and without sampling techniques. By contrasting the influence of sample techniques on prediction performance, it is possible to assess the impact of sampling procedures on prediction performance. Table 2.1 and Table 2.2. For all four classifiers, the accuracy of the classifier taught without sampling methods was greater than that of the classifier trained using sampling approaches. This, however, does not rule out the possibility of success by following a model approach. Classifiers are more likely to predict the 'on-time' class if they are trained on data that is unbalanced. Flight delays are more likely to be considered on-time than flights that arrive on time.

It illustrates that deep belief networks and neural networks can be used to anticipate flight delays. The accuracy of prediction using neural networks is 92 percent. with three neurons in one buried layer The use of deep nets has become increasingly popular. With four neurons in each of two groups, the accuracy was 77 percent. layers that are not visible This work makes a significant contribution in that it confirms the efficacy of neural networks and other comparable techniques Deep architectures are used to classify aircraft delays and no-delays. The accuracy of the data is currently being investigated. Deep Nets are used to make predictions. Application will be a part of future work. A variety of deep architectures to compare and contrast. The algorithm is the most effective. When this is completed, a client will be contacted. It is possible to create a focused application that takes attributes. This is something that the passenger is aware of. This could be a huge step forward in order to alleviate the annoyances that airplane customers endure.

Figure 2.8 depicts the four classifiers' ROC curves and AUC values. They provide evidence to support the claim that, out of the four classifiers, Random forest has the strongest capacity to differentiate between planes that are late and those that arrive on time. mainly due to the fact that its AUC is the highest and

Classifier	Accuracy (%)	Time
Random Forest	83.40	9
AdaBoost	83.21	12
kNN	82.42	2
Decision Trees	82.84	0

Table 2.2: Accuracy of training and time

it is situated the farthest away from the diagonal line. Nevertheless, the trend of the ROC curves does not always correlate to the accuracy of the prediction. The k-nearest-neighbors classifier has the lowest area under the receiver operating characteristic curve (AUC), however the k-nearest-neighbors classifier also has the lowest accuracy (a). In the confusion matrix, the k-nearest-neighbors had a high TPR while also having a high FPR. On the other side, the FPR and TPR are both reduced when using the decision trees classifier. It's possible that the costs of false positives and false negatives may be utilized to decide which classification method is the most effective. A classifier with a high true positive rate is preferable in situations when the cost of false negatives is considerable. [16]

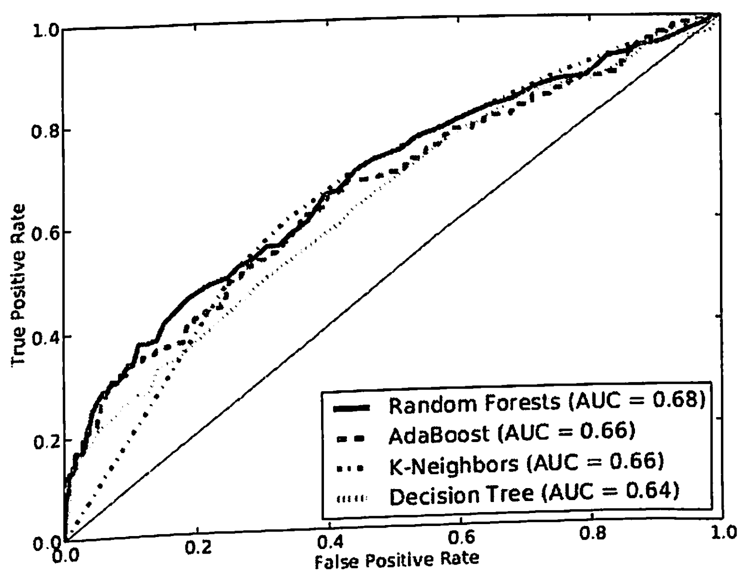


Figure 2.8: Receiver Operating Characteristic of Classifiers

2.2.3 A Machine Learning and Deep Learning Iterative Approach

Data collection is critical since it serves as the foundation for training the model. (Figure 2.9) Because of the limited resources for processing, the data is preprocessed using stratified sampling to build a smâller dataset from the current huge dataset that properly reflects the complete population. The preprocessed data is then utilized for feature extraction, which extracts key properties before training the neural network model and deep belief nets using the training data set. The models are repeatedly trained to produce the best results. The models are then verified for accuracy using a testing data set. In addition, additional unclassified cases are fed into the classifier in real time to anticipate whether a flight would be delayed or not.

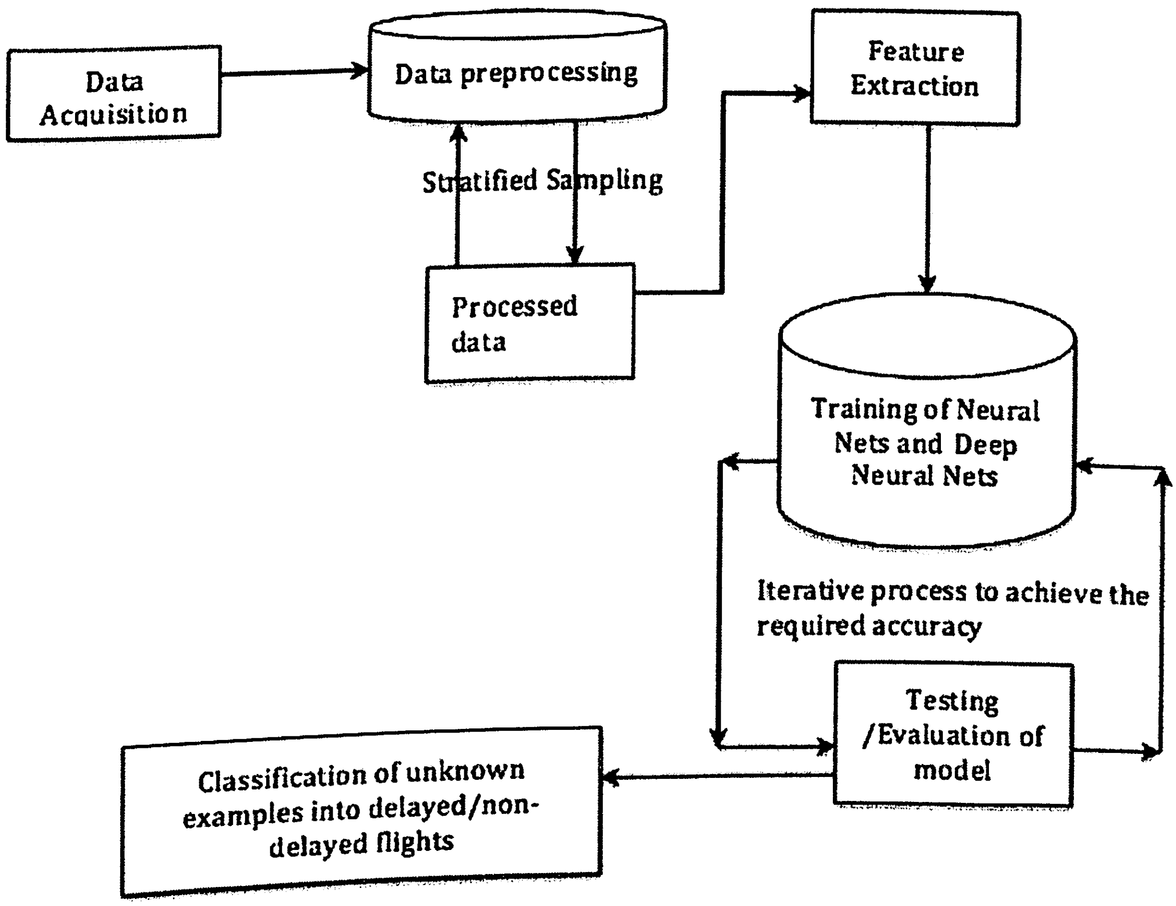


Figure 2.9: The framework for prediction of delay

The machine learning-based categorization models studied for delay prediction in this article are ANN and DBN. Both classifier models are developed using a stratified random sample of 8000 training instances collected from the real dataset.

The 6 derived qualities are used as inputs to each of these classifiers, and the intended outcome is binary, with 1 representing DELAY. NO DELAY is represented by a value of 0. Artificial Neural Networks are made up of nodes (input, hidden, and output) that have activation functions applied to them. The robust back propagation [32] learning algorithm is used to train ANNs, which is similar to the back propagation technique. However, robust back propagation has the benefit of training the neural network quicker than back propagation. Because it is a first-order learning method, modest changes in parameter values have no effect on the program's convergence time. It is more stable in terms of learning rate and momentum. [35]

When it comes to DBN (Deep Belief Network), it's usually a stacked arrangement of Restricted Boltzmann Machines (RBM). DBN consists of a sequence of convolutional layers, each of which admits only connections from the one before it. It turns the neurons in the input layer into a feature vector that can be classified by many connected layers. All parameters are fine-tuned to reduce the misclassification error throughout the training data set by fine-tuning learning rate and momentum. The contrastive divergence technique is used to pre-train RBMs one at a time. This pre-training aids in the initialization of the network's weights. The Deep Net is then fine-tuned using the back propagation mechanism.

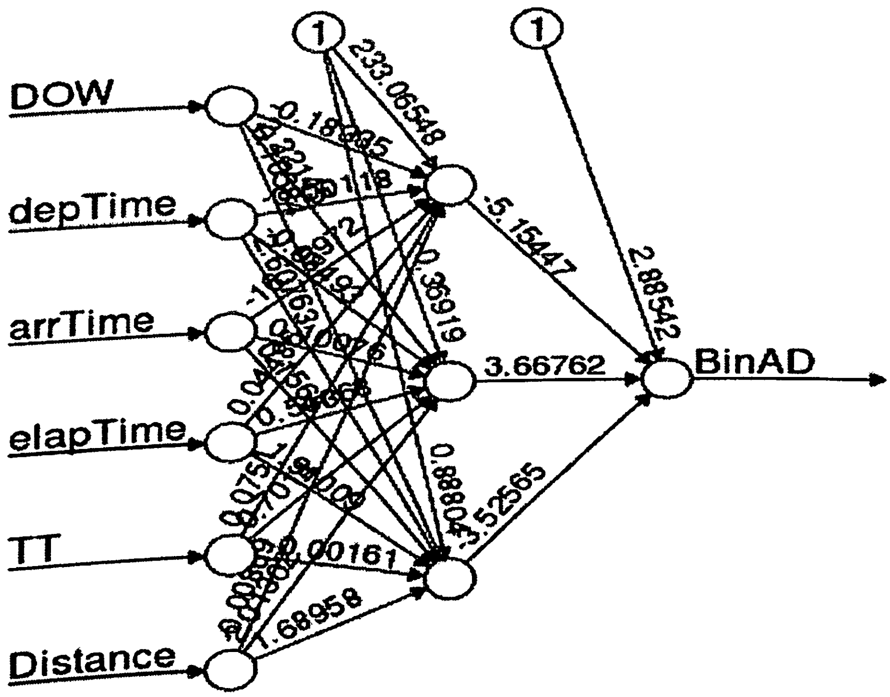


Figure 2.10: Neural network plot for the dataset

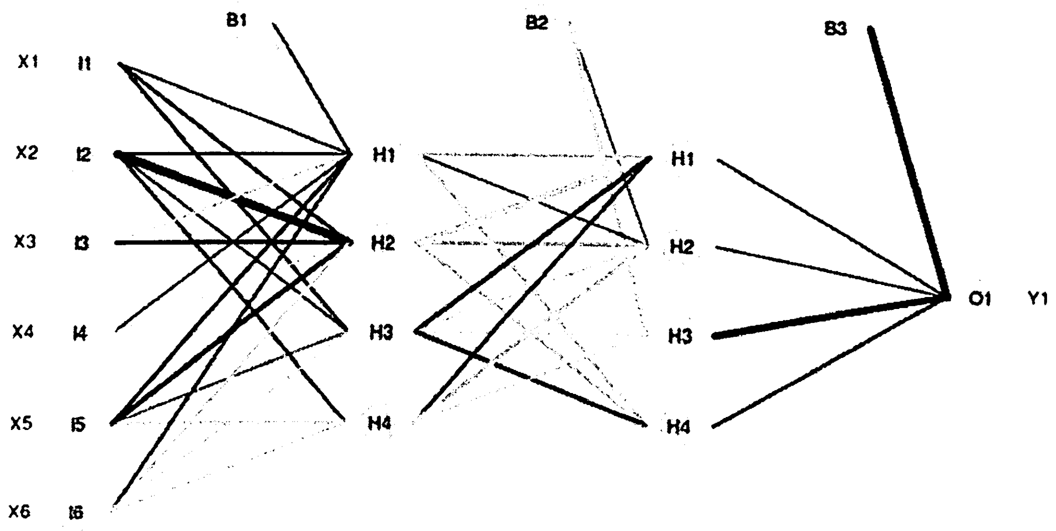


Figure 2.11: Deep Belief network plot for the dataset

Figures 2.10 and 2.11 depict these two classifier models, with numbers on each node denoting the weights supplied by the algorithm to the nodes. RStudio was used to create this diagram. These weights are seen to fluctuate each time this training occurs. Bias is another variable that is always changing. Bias is a number that helps alter the activation function by a certain amount. The classifier model is trained indefinitely until the network converges.

2.3 Related works

Mueller and Chatterji used a statistical approach to analyze departure and arrival data, as well as to describe the delay data. [27] Statistics have also been utilized by Tu et al. citefifthone to construct an estimate of the long- and short-term trends in aircraft departure delay distributions. Because statistical approaches depend on estimates and approximations, the findings they provide may lead to incorrect conclusions. [34]

Individual flights have been the subject of most research on flight delay prediction. Based on previous flight data, Assent et al. [4] divide flights into three categories: "ahead of time," "on time," and "delayed," with a stated accuracy of up to 45.4 percent. The authors tweaked their classifier to account for regionally variable relevance characteristics. The classifier, on the other hand, ignores information about the surroundings, which will be addressed in future research. Belcastro et al. [9] use flight and meteorological data to anticipate individual aircraft delays based on different delay criteria. The tolerance for deviation from the intended arrival time in order to be deemed on time is determined by the threshold. As a consequence, by combining meteorological data from the origin and destination airports, the classification findings attain an accuracy of up to 85.8% for a 60-minute delay threshold. Lowering the threshold reduces the accuracy of categorization.

To anticipate aggregated air traffic delays, Rebollo et al. [29] simulate the US airport network. As a result, forecasting necessitates knowledge of the whole airport network. The inputs are the global delay state of the National Airspace System and the delay states of the most significant airports. Meteorological data or any other information about the state of the environment is not utilized to make predictions. Prediction accuracy is 81 percent for a delay threshold of 60 minutes.

Choi et al. [16] use meteorological data at the origin and destination airports to identify delays of particular pairings of origin and destination airports. The purpose is to estimate whether an aircraft will arrive on schedule or will be delayed. A flight is deemed on time if it is delayed for less than 15 minutes. Random forest, AdaBoost, KNN, and decision trees are among the classifiers tested. The random forest classifier outperformed the others as a consequence. Yi et al. [19] use multiple linear regression on historical flight data that includes certain meteorological variables, such as wind direction and speed, to estimate flight delay. The purpose is to anticipate if an aircraft will be delayed for more than 30 minutes. By reaching 79.1 percent accuracy, the linear regression model surpassed both bayes and C4.5. According to the current state of knowledge, weather circumstances have a critical role in the forecast of delay. According to Abdelghany et al. [1], weather is responsible for roughly 75% of delays caused by tight connections, and this number is expected to rise if no action is done. It has been researched for years, and the modeling methodologies may be divided into three categories. The first category is statistical and probability analysis. Historical data was used to examine the factors that influence flight delays and delay propagation characteristics. There have been attempts to predict the probability distribution of flight delays in certain studies. Despite extensive research, it remained difficult to estimate the delay of a certain aircraft. Modeling and simulation approaches [7] fall within the second group. The real flight operation procedure was imitated by creating a model in the simulation system. However, the simulation speed is rather slow, which makes it ineffective for large-scale flights. Using machine learning, a third option, allows researchers to take use of existing data while also predicting future data. It is possible to make use of machine learning methods such as support vector machines, such as SVMs, Random Forests, and K-Nearest Neighbor (KNN). Nonlinear issues, such as predicting aircraft delays, are especially well-suited to SVM. However, there is still room for growth. To begin, improving the stated feature variables may help increase the model's generalization capabilities. Flight plans are the same every day because of the regularity of flight schedules, but the length of time it takes to go from point A to point B is vastly different. Such tag-hopping may result in model failures or problems. Flow control, weather, and the quantity of aircraft are all unexpected, even if the daily average is fairly consistent. Some models fail to account for hourly flight delays in historical data, which should be taken into account when determining delayed patterns for particular trips. Reduced ceiling and visibility are the biggest drivers to substantial delays, according to Cruciol et al. [18] found that air temperature and wind speed are related to air traffic delay when using pattern recognition in air traffic flow management. Furthermore, the classifier proposed by Belcastro et al. [10] for predicting flight delays based on historical flight data and meteorological data, which includes the two attributes ceiling shows that good prediction results can be achieved using these data.

Researchers developed analytical methods that assisted them in extracting characteristics from the model using approaches such as Natural Language Processing, Nave Bayes, and Support Vector Machine. The vast majority of them were concerned with predicting overall flight delays. The goal of our study was to anticipate flight delays for a particular airport over a certain length of time. We looked at the importance of each feature using a regression model, and then used a feature selection technique to look at the effect of combining features. These two strategies were used to determine which qualities should be retained in the model. Instead of testing several machine learning models on the complete set, we sampled 5,000 records at a time. The machine learning models employed in this work were the Random Forest classifier and the Support Vector Machine classifier. In addition, we used the One-Hot-Encoder technique 5 to develop a variation of the model for testing prospective prediction performance.

by creating a model in the simulation system. However, the simulation speed is rather slow, which makes it ineffective for large-scale flights. Using machine learning, a third option, allows researchers to take use of existing data while also predicting future data. It is possible to make use of machine learning methods such as support vector machines, such as SVMs, Random Forests, and K-Nearest Neighbor (KNN). Nonlinear issues, such as predicting aircraft delays, are especially well-suited to SVM. However, there is still room for growth. To begin, improving the stated feature variables may help increase the model's generalization capabilities. Flight plans are the same every day because of the regularity of flight schedules, but the length of time it takes to go from point A to point B is vastly different. Such tag-hopping may result in model failures or problems. Flow control, weather, and the quantity of aircraft are all unexpected, even if the daily average is fairly consistent. Some models fail to account for hourly flight delays in historical data, which should be taken into account when determining delayed patterns for particular trips. Reduced ceiling and visibility are the biggest drivers to substantial delays, according to Cruciol et al. [18] found that air temperature and wind speed are related to air traffic delay when using pattern recognition in air traffic flow management. Furthermore, the classifier proposed by Belcastro et al. [10] for predicting flight delays based on historical flight data and meteorological data, which includes the two attributes ceiling shows that good prediction results can be achieved using these data.

Researchers developed analytical methods that assisted them in extracting characteristics from the model using approaches such as Natural Language Processing, Nave Bayes, and Support Vector Machine. The vast majority of them were concerned with predicting overall flight delays. The goal of our study was to anticipate flight delays for a particular airport over a certain length of time. We looked at the importance of each feature using a regression model, and then used a feature selection technique to look at the effect of combining features. These two strategies were used to determine which qualities should be retained in the model. Instead of testing several machine learning models on the complete set, we sampled 5,000 records at a time. The machine learning models employed in this work were the Random Forest classifier and the Support Vector Machine classifier. In addition, we used the One-Hot-Encoder technique 5 to develop a variation of the model for testing prospective prediction performance.

Various machine learning approaches have been used by many academics to investigate the topic of airplane delays. Esmaeilzadeh and Mokhtarimousavi used a support vector machine to mine the nonlinear relationship between flight delay and a variety of variables. Because of the black-box nature of machine learning, a sensitivity analysis of related and independent variables was conducted, taking into consideration weather, airport scene operation, demand, and other factors. This research [3] provided a fresh viewpoint on the causes of flight delays. An XGBoost-based aircraft arrival delay prediction classification model and a linear regression-based flight arrival delay prediction regression model were proposed by Kalyani et al. XGBoost is a Decision Tree-based ensemble learning method that can discover the optimum result by continuously adjusting the hyperparameters. Because of its fundamental assumption and ease of implementation, linear regression is one of the most often used algorithms in machine learning. With a prediction accuracy of 0.77, Zhang and Ma created a Catboost algorithm-based flight delay prediction model. The SHAP value [5] was used to determine the contribution degree of the qualities. Khaksar and Sheikholeslami developed a hybrid strategy that used the J48 Decision Tree and K-means to train flight datasets from the US and Iran, respectively, and compared it to four algorithms before concluding that the hybrid method generated the best results.

Most researchers would use several machine learning algorithms to train the same datasets and compare evaluation indexes to determine the best algorithm and best forecast outcome when utilizing machine learning methods. Furthermore, the number of algorithms accessible expands as machine learning technology progresses, and most researchers use at least three algorithms in a single study. Based on Hartsfield-Jackson International Airport, Henriques and Feiteira suggested a decision tree, random forest, and multilayer perceptron classification model. The Multilayer Perceptron provides the highest level of precision. Choi et al. tested two supervised learning algorithms, Decision Tree and KNN, as well as two ensemble learning algorithms, Random Forest and Adaboost, and discovered that the ensemble algorithm classifier performed better than the single algorithm classifier [10]. Stefanovi et al. investigated the Lithuania Airport flight delays datasets and discovered that seven machine learning algorithms, including probabilistic neural networks, multilayer perceptron neural networks, Gradient-Boosted Tree, Decision Tree, and Gradient-Boosted Tree, outperformed the others. The research efforts stated above are inspirational, and the majority of them employ model comparison to discover one optimal model while ignoring the others, squandering processing resources. Flight datasets are also vast and versatile, and method stability is essential for real-world applications. However, most studies, particularly those that used novel algorithms, overlooked algorithm stability. In this research, we develop a Stacking-based flight delay prediction classification model and conduct tests to establish Stacking's stability.

Flight delay prediction systems based on machine learning are rapidly improving. Feature selection, on the other hand, is an important method that has been missed in previous studies [12]. In machine learning, selecting features is a critical step [13]. The primary purpose of feature selection is to reduce duplicate features and improve model efficiency by assessing feature relevance. Onan and Korukoglu proposed a feature selection model based on the ensemble approach. The experiment's findings show that the recommended approach not only effectively handled difficult data, but significantly improved classification accuracy. Using weather data to improve prediction accuracy is also a possibility, albeit reliable weather data may not be available until a few hours before the flight.

Ai Pan et al. employed a convolutional LSTM (Conv-LSTM) method to investigate the flight delay distribution, taking into account indirect factors such pre-order flight delay, route congestion, and airport capacity [14]. The study employed a limited number of parameters due to the enormous amount of data and the wide range of predicted time windows. Hao et al. utilized historical data from several airports to develop a multi-step prediction of airport delay using spatio-temporal data. The model has a good forecast accuracy, however it cannot analyze a single trip. Yu et al. used a deep belief network and SVR to mine the contributing elements of a flight delay, and the results were better than benchmark approaches like k-nearest neighbors (KNN), support vector machine (SVM), and linear regression (LR). The report, on the other hand, relied on a statistical approach to analyze crucial aspects and was unable to provide a more intuitive, thorough, tactical account of each delay. Flight delay analysis and the identification of relevant factors are getting increasingly comprehensive. The influence of convective weather, local weather, and airport traffic demand on ground delay was studied by Liu et al. using machine learning algorithms such as SVM, LR, and RF. They couldn't, however, direct specific projects. From two classifications to multi-classification, research on delay prediction is evolving. To categorize and regress airport delays, Gui and Liu used a combination of machine learning and deep learning algorithms, as well as data on weather, planes, airports, and other direct affecting variables. To estimate aircraft delays at Boston Logan international airport, Jia Honghai et al. used KNN, RF, LG, Decision Tree, and Gaussian Nave Bayes with directed factors. The results revealed that the stochastic forest model outperformed baseline methods in terms of prediction accuracy, however no variables influencing prediction outcomes were examined. Khan et al. used machine learning techniques to investigate the delay and its duration, however they did not take into account indirect effects. Junfeng et al. constructed eight commonly used modes, including linear regression models, non-linear regression models, and tree-based ensemble models , and discovered that if the feature set could capture the arrival characteristics, even simple linear regression models or algorithms could complete the prediction task. The DL technique based on multi-airport data has gotten a lot of interest as computational power has increased. Higher data dimensions, such as oD (origin-destination) data between airports, are required for multi-airport delay prediction, which will not allow the model to be updated. at the same time, giving a detailed analysis for a single trip is challenging.

Short-term delay prediction (for the next several hours or so) is a well-developed area. Indeed, combining data on weather, airport congestion, and existing flight delays, it is possible to make quite accurate forecasts of future delays, since some of the factors that influence. They are well-known (even if they still have a random component). The website FlightCaster, for example, analyzes a variety of data sources (airports, airlines, weather, and maybe previous data) to provide travelers with the likelihood of arriving on time, less than one hour late, or more than one hour late. When no short-term information is available, however, this website uses the same estimates for all flights. Many studies on the management and transmission of aircraft delays have been undertaken, with an emphasis on traffic management systems. Which might be useful in traffic management systems. These early models do not take any flying peculiarities into consideration and just show global trends. Another study concentrating on Ground Delay Programs Improvement by Allan et al looks into the meteorological conditions and their impact on on-ground and aircraft delays. Zonglei et al [3] employ decision trees and neural networks to forecast an airport's total traffic status in the near future. Finally, Tu and Ball sought to explain the underlying processes in [4] by modeling three components: a seasonal trend, a daily trend, and a random residual, all of which were fitted using a genetic algorithm. Their method seems to take a long time to compute and hasn't been extensively tested. We will, however, strive to compare our future results to those expressed in this essay since it is close to my project's goal.

3. Research method

3.1 The Linear Regression

One or more explanatory variables (or independent variables) denoted X in statistics are used to describe the relationship between a scalar dependent variable y and the dependant y. When just one explanatory variable is present, simple linear regression is the method of choice. Multiple linear regression is used when there are a large number of variables. As a result, linear predictor functions, whose parameters are inferred from the data using linear regression, are used to model the connections. Models that fall under this category include linear ones. Linear regression has a wide range of uses. There are two broad groups of applications: (1) If the goal is prediction, forecasting, or error reduction, linear regression may be used to fit a predictive model to a collection of observed Y and X values. The fitted model may be used to predict the value of Y if an additional value of X is given without its corresponding value of Y after such a model has been constructed. Second, linear regression analysis may be used to quantify the strength of the link between Y and the Xj and to discover which Xj may have no association with Y at all, as well as which subsets of the Xj contain redundant information about the Y variable.

Teh study and design of complex and large-scale systems with numerous variables necessitates the development of novel methods for identifying, classifying, and analyzing massive amounts of data. As a result, academics have proposed data mining ways for detecting, gathering, categorizing, and sorting useful information from databases, as well as producing and storing it. Machine learning algorithms are used in these techniques to gain information from a databank. Data mining may also be used to anticipate what will happen in the future.

A decision tree's usual structure includes a root node, numerous branches, and a large number of leaf nodes. In this method, the user gradually builds a decision tree based on the dataset being divided into mini subsets until he or she reaches a tree containing decision nodes and the nodes of the leaves. All of the nodes in this tree indicate attributes that are being tested, and the branches reflect the results of those tests. The tree's highest point is the root node. Figure 3.1 depicts a typical decision tree relating to airline flight delays. A scheduled departure is the tree's root, while other characteristics like fleet age and visibility distance are its leaves, as shown in the diagram. The eet is roughly 25 years old on average. Despite the fact that this age is higher than the global average, it is the average age recorded in our database due to fines imposed in past years.

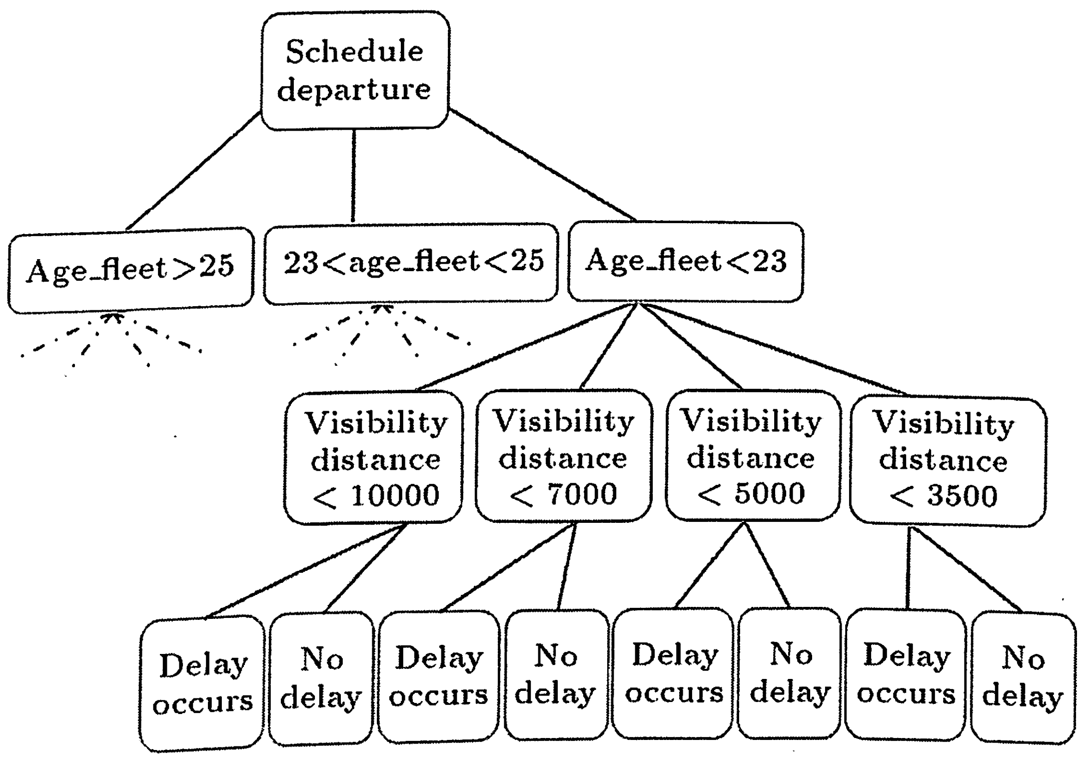


Figure 3.1: A typical decision tree for delay prediction

Cluster analysis is a widely used and significant data analysis approach. It may be used to describe a dataset, and then it can be analyzed using a decision tree. We employ a hybrid technique, which is a decision tree based on the clustering algorithm, to enhance the outcomes of standalone approaches. Clustering is used as a down-sampling preprocess for classification in this approach to minimize the size of a training set. As a consequence, the dimensionality of the issue is decreased, and the classification challenge is smaller and simpler to answer. As a result, a decision tree based on the premise that each cluster belongs to a class is constructed using K-means clustering for classification. In this study, such classification and assumptions are used. Figure 3.2 depicts the hybrid approach's underlying mechanism. To begin, data consumption and interaction are preprocessed. Then, depending on whether you want to pick just a subset of attributes variables or use all of them, you may use an optional attribute selection method. Finally, training data is used to run a clustering algorithm. A crucial condition is that the number of created clusters matches the number of class labels in the dataset. This similarity enables the development of an usable model that links each cluster to a single class.

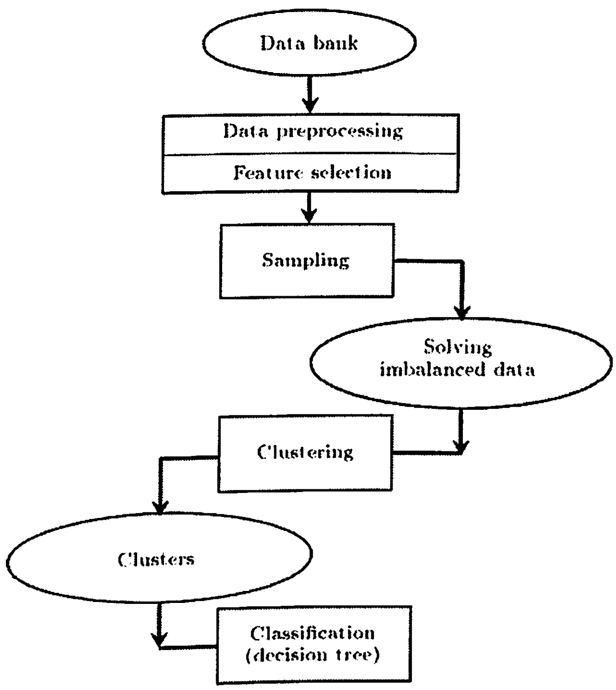


Figure 3.2: Process of hybrid approach

The confusion matrix may be used to calculate the classifier's performance. The classifier results may yield four values as compared to the actual outcome:

True Positive indicates that the anticipated value is positive and that the actual value is also positive. True Negative is the projected value is minus, as is the actual value. The anticipated value is positive, while the actual value is negative. False Positive (FP) The anticipated value is negative, while the actual value is positive. False Negative (FN): the expected value is negative, but the actual value is positive. The performance of chosen algorithms was assessed using four metrics: accuracy, precision, recall, and F1 score. These variables are all linked to algorithm quality in a good way. As a result, the higher the values of these measurements are for a certain algorithm, the better it performs. Calculations employing these parameters may be used to determine the value of the four measures:

$$Accuracy = \frac{TP + TN}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

Figure 3.3: The confusion matrix formulas description

Rather than forecasting the complete delay probability distribution, we could be more interested in a simpler predictor that yields probabilities of delay or no delay (we will define the boundary between the two classes later). A binary classifier is a predictor that makes a prediction based on a collection of parameters associated with an observation. It is considered to be discrete when a binary classifier only returns True (in our case with delay) or False (without delay). In this case, the classifier's performance can easily be assessed by comparing the returned classes to the test samples' classes using a confusion matrix and associated metrics (specificity, sensitivity, accuracy, etc.). A good classifier should have a high true response rate (the ratio of correctly predicted delays to the total number of actual delays; a TPR of 1 indicates that all delays were correctly predicted) and a low false response rate (which is the ratio of timely samples predicted as delayed to the total number of timely samples; FPR equal to 0 indicates that no timely sample was predicted as delayed).

Some binary classifiers are continuous, yielding the odds of executing delayed

P (delayed) and not running delayed P (not delayed) (not delayed). As previously stated, a discrete classifier may be formed by specifying a probability threshold pthreshold, so that if P(delayed) > pthreshold, the classifier returns "delayed," else "no delay." As a consequence, each pthreshold value produces a distinct discrete classifier with a distinct performance.

The receiver performance curve (also known as the ROC curve) shows the performance of all discrete classifiers created from a continuous classifier. It reflects the ratio of true positives to false positives for each classifier (see Figure 2.5).

The reverse classifier (which delivers the opposite answer) will perform better if the TPR is less than the FPR. If TPR and FPR (diagonal line) are equal, the classifier is similar to the random assumption. Finally, there will be an ideal classifier (TPR = 1 and FPR = 0) in the upper left corner.

As a consequence, we can measure the classifier's performance by calculating the area under the ROC curve to see how near it can work to the top left corner (abbreviated as AUC).

According to the preceding explanation, AUC for a random classifier is 0.5 and for an ideal classifier is 1. AUC also facilitates evaluating the performance of many classifiers simpler than comparing ROC curves. The AUC of the ROC curve in Figure 2.5, for example, is 0.62.

3.2 XGBoost

XGBoost (Extreme Gradient Boosting) is a scalable machine learning platform with a distributed decision tree for gradient boosting (GBDT). It contains parallel tree boosting and is the best machine learning tool for regression, classification and ranking tasks.

Understanding the machine learning principles and approaches that XGBoost is founded on is the first step to understanding the software's capabilities.

Algorithms trained on a dataset containing labels and features are then used to predict additional features in the dataset based on the learnt model's predictions.

Gradient boosting is a method of 'strengthening' or 'improving' one weak model by combining it with several other weak models to form a strong model in the aggregate. In gradient boosting, which is an extension of boosting, the approach to building an additive weak model is defined as a gradient descent algorithm for the target function. Increasing the gradient provides the expected results for the next model to minimise the errors. The expected results for each instance are determined by the error gradient (hence the name gradient enhancement) with respect to the prediction.

In recent years, XGBoost has gained a lot of attention for its ability to assist individuals and teams in winning practically all Kaggle structured data challenges. Companies and academic organizations submit data to these contests, and statisticians and data collectors compete to create the best data prediction and comprehension models.

The library has a minimum of frills, as it focuses entirely on computational speed and model performance. However, it provides many sophisticated features. The model incorporates parts from scikit-learn and R implementations, as well as new features like regularization. Gradient amplification may be achieved in three ways: 1. The gradient enhancement algorithm, also known as the gradient enhancement machine, incorporates speed learning. 2. Subsampling on various levels using stochastic gradient improvement in the row, column, and column 3. L1 and L2 regularized regularized gradient boosting.

3.3 Recurrent Neural Network

Using distinct predictors for each combination of category traits is the simplest method to deal with them. The unique properties of each combination are thereby discovered, rather than being lost amid the rest of the data. As a consequence, we have less data points with which to train the predictor.

The departure point, destination, and airline are categorical elements that we estimate for each journey, resulting in a unique route (referred to as a "plot"). Roughly 8,600 routes are included in the 2010 data, with around 7,250 having more than 50 yearly entries (i.e. at least one flight per week). Airports have their own unique features, such as traffic patterns and typical weather patterns. The density of airline timetables, the quality of service, and the accuracy of flight length projections are all unique features.

A kernel density estimate or a kernel conditional density estimate (with limitations on the hour of arrival, day of the week, and month) may be used as a predictor for any combination of variables. The option of whether to optimize capacity for each predictor individually or to utilize the same capacity for all predictors is one issue with this strategy.

According to studies, the combination of origin and destination airports (OD) can have a significant impact on flight delays. In addition to the three databases, airport components (BTS, LCD, ASPM) were included to study the impact of linked airports. In the OD pair feature, 30 airport features and one directional feature are connected to ORD through a network of airports. Directional features are turned on or off depending on whether you're starting from or concluding your journey at ORD. When flying from ORD to LGA, the LGA function is one and all the other 29 airport functions are 0, e.g. The total number of measurements is reduced from 60 to 31 using the destination function.

Some researchers use a statistical method to study airline delays using probability models. Muller and Chatterjee used a Poisson distribution and a normal distribution to model the probability density function of flight delay times. To estimate the impact of airport congestion management techniques, Avijit et al. developed models of average delays and probability of cancellation. The probability of flight delay was calculated by Wesonga et al. based on multiple factors, including aircraft type, number of passengers, and meteorological conditions.

To explain the results of several regression and classification models, Juan José Rebollo and Hamsa Balakrishnan used hundreds of pairs of departure and destination points. The results show that, of all the approaches studied, the random forest outperformed them all. Variables such as number of origin and destination pairs and forecasting horizon affect predictability. Multiple linear regression was used by Shruti Oza and Somya Sharma to predict weather-induced flight delays in flight data, as well as climate parameters and the probability of delays due to weather conditions. The forecasts were based on a number of key factors, including airline, departure time, arrival time, departure point and destination.

LSTM is a popular RNN category that was first introduced in 1997. Text, video and audio processing has benefited from RNN, and it is gradually finding application in additional disciplines. However, when the input sequence data is long enough, preserving previous information far from the decision point is a challenge for a typical RNN. LSTM has a natural advantage when dealing with long term dependency of sequential data. The incoming information is filtered in each LSTM block, while the previous historical information is absorbed to preserve the long-term memory of the sequence data. internal links between sequences are well preserved with LSTM, which contributes to the retention of deleted information. it is assumed that the airport will be able to estimate the number of aircraft that will arrive next. as a result, the input data for our delay prediction model should be sequence data, and the output data should also be sequence data. However, structural limitations make sequential data prediction problematic for LSTM.

Cho et al. proposed the seq2seq approach, which uses sequence data as input and output data. Seq2seq is a deep learning technique with important applications such as machine translation, document extraction and human-machine interaction, and its input and output characteristics are comparable to those of delay prediction in this paper. The encoder and decoder are included in the Seq2seq paradigm. The length of the encoder is fixed, and the past data it uses data from the previous 50 flights into a vector and then used the vector to predict the next 10 delays one by one. They pay attention to the same parts of the vector because they pay attention to the same parts of the vector. With the seq2seq paradigm alone, it is difficult to achieve the desired result.

To compensate for the lack of information in the seq2seq model, Bahdanau and Cho et al. included an attention mechanism. Seq2seq with an attention mechanism examines the context vector and various values of prior information at the decoder level. Data with input and output sequences can be described without input and output distances after adding an attention mechanism.

To anticipate individual flight delays, random forest and LSTM based architectures were used. Experimental results show that the random forest method can work well in binary classification tasks, but there is still potential to improve multi-category classification tasks. The LSTM based architecture has great learning accuracy, indicating that the LSTM cell is a good structure for dealing with temporal sequences. On the other hand, the problem of overfitting in an LSTM-based design has not yet been solved. In conclusion, when dealing with a limited dataset, the random forest based scheme provided excellent adaptability at the expense of learning accuracy. Our future work will focus on collecting or creating more data for training, integrating more information such as airport traffic flow and visibility into our dataset, and developing more delicate networks to reduce overfitting and increase more category clustering task test accuracy.

4. Results and Discussions

Flight schedules available at Almaty International Airport (ALA) from Air Astana airline company between January 1, 2019 and April 30, 2022 are used in this study.

4.1 Results and Discussions

Dep	Avg. delay minutes	Departure
ALA	29.853850	6767
NQZ	30.739626	2458
TSE	30.501841	1901
GUW	29.958185	1124
CIT	31.073937	1082
SCO	30.497722	878
AKX	30.831541	558
URA	31.136449	535
DXB	28.822857	525
IST	29.020833	480

Figure 4.1: Average delay minutes by departure

The data set was tweaked somewhat, including altering the variable "DEW POINT" from object to integer datatype since it contains numerical values and removing rows with null values. There were just two missing values out of 28820 rows, indicating that the deletion will have minimal impact on the overall distribution of the data set.

Some of the algorithms used can only work with numerical data. The data collection includes both numerical and category variables, as seen in the preceding

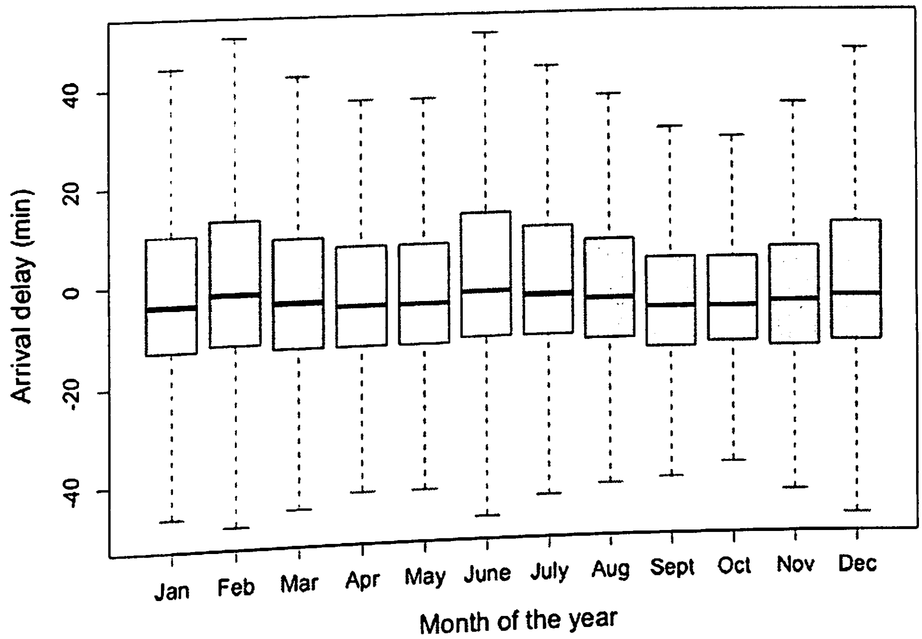


Figure 4.2: Average delay minutes by departure

Type	Afternoon	Early Morning	Evening	Morning	Night	Noon
DepDelay						
False	11646	1386	6906	13491	6224	8338
True	2947	104	2456	1646	2126	1671

Figure 4.3: Maximum number of flight operations performed between 07:00AM - 11:00AM

section. To prevent these algorithms failing when dealing with categorical data, these variables must be transformed to numerical variables. Integer encoding, which converts category labels to unique integer values, was utilized for both training and fresh data [31]. Also deleted were category factors like "TAIL NUM," which have minimal impact on forecasting flight delays.

A supervised machine learning technique was used in this research. The data set contains a target variable, and the purpose is usually to teach the computer a categorization system that has been established [5]. The major goal of this research is to use data from labels to anticipate flight delays. As a result, a supervised learning classification method was chosen as the best fit. The prediction of flight delays was seen as a binary classification issue, in which supplied data was used to predict whether or not a flight delay will occur. Following talks with ex-

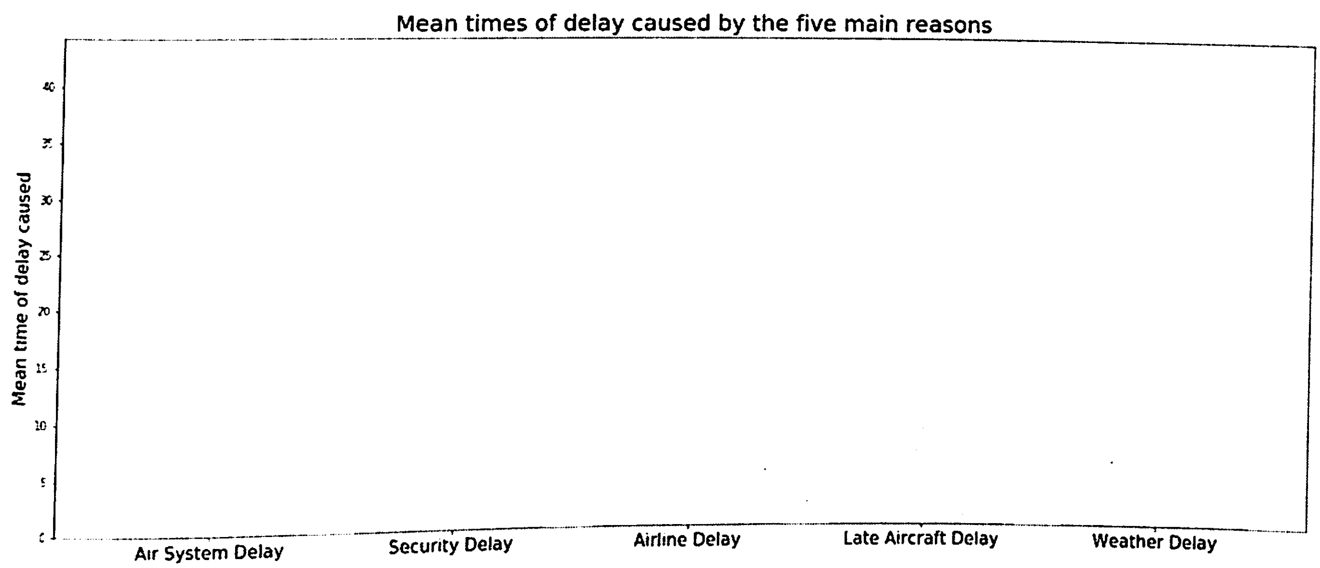


Figure 4.4: Mean times of delay caused by the five main reasons

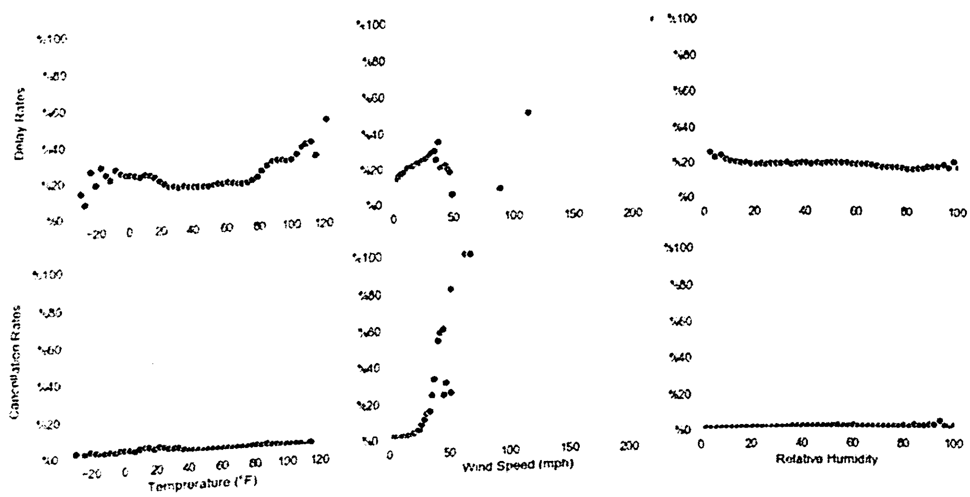


Figure 4.5: Effect of weather conditions on delays and cancellations

perts and past work in the airline area, the definition was established that a flight is delayed if the variable "DEP DELAY" (minute difference between planned departure time and actual departure time) is larger than 15. Otherwise, it will not be delayed. As a consequence, a new binary variable "IS DELAY" was generated, with the value 1 for a delayed flight and 0 for a non-delayed trip.

The data set contains 3873 delayed flights and 24945 non-delayed flights, indicating an uneven distribution since the majority of flights were not delayed, according to the variable "IS DELAY." This issue was solved using 10-fold cross-validation. It was responsible for the development of the training and testing sets. On the testing set, each algorithm was run with the scikit-learn python package's default parameters, and the training set was the same for all methods.

Due to the unbalanced structure of the data set, each method has to apply weighted precision, recall, and F1 score. For each label, the ratio of accurately predicted samples is determined first. The weighted average is calculated by multiplying these ratios by their percentage to the total number of samples.

5. Conclusion

An effective and realistic flight delay prediction system would result from combining these parameters in the right proportions. The analysis could be effectively extended in the future to include the development of multiple models utilizing machine learning and deep learning approaches. Modeling methodologies might benefit from a thorough examination of numerous trends and patterns among diverse components, as well as their changing interactions with one another and through time. Data mining and gathering would provide a stronger foundation for statistical models. The optimal performance of classifiers may be readily identified when the expenses of pros and cons considered. It might thus serve as a good basis for a controlling instrument that predicts plane arrival. A comprehensive examination of prediction uncertainty would also improve the model's predictive performance.

References

[1] Khaled Abdelghany et al. "A model for projecting flight delays during irregular operation conditions". In: *Journal of Air Transport Management - J AIR TRANSP MANAG* 10 (Nov. 2004), pp. 385–394. DOI: 10.1016/j.jairtraman.2004.06.008.

[2] Kerem Akartunalı et al. "Airline planning benchmark problems—Part II: Passenger groups, utility and demand allocation". In: *Computers Operations Research* 40 (Mar. 2012). DOI: 10.1016/j.cor.2012.03.005.

[3] Hajar Alla, Lahcen Moumoun, and Youssef Balouki. "A Multilayer Perceptron Neural Network with Selective-Data Training for Flight Arrival Delay Prediction". In: *Scientific Programming* 2021 (June 2021), pp. 1–12. DOI: 10.1155/2021/5558918.

[4] Ira Assent et al. "Data Mining For Robust Flight Scheduling". In: Jan. 2009, pp. 267–282. ISBN: 978-0-387-79419-8. DOI: 10.1007/978-0-387-79420-4_19.

[5] Taiwo Ayodele. "Types of Machine Learning Algorithms". In: Feb. 2010. ISBN: 978-953-307-034-6. DOI: 10.5772/9385.

[6] Michael Ball et al. "Total Delay Impact Study: A Comprehensive Assessment of the Costs and Impacts of Flight Delay in the United States". In: (Oct. 2010).

[7] Cynthia Barnhart and Amy Cohn. "Airline Schedule Planning: Accomplishments and Opportunities". In: *Manufacturing Service Operations Management* 6 (Dec. 2004), pp. 3–22. DOI: 10.1287/msom.1030.0018.

[8] Massoud Bazargan. *Airline Operations and Scheduling*. Mar. 2016. ISBN: 9781317182924. DOI: 10.4324/9781315566474.

[9] L. Belcastro et al. "Using Scalable Data Mining for Predicting Flight Delays". In: *ACM Transactions on Intelligent Systems and Technology* 8 (Jan. 2016). DOI: 10.1145/2888402.

[10] L. Belcastro et al. "Using Scalable Data Mining for Predicting Flight Delays". In: *ACM Transactions on Intelligent Systems and Technology* 8 (Jan. 2016). DOI: 10.1145/2888402.

[11] Reshma Bhagat and Sachin Patil. "Enhanced SMOTE algorithm for classification of imbalanced big-data using Random Forest". In: *Souvenir of the 2015 IEEE International Advance Computing Conference, IACC 2015* (July 2015), pp. 403–408. DOI: 10.1109/IADCC.2015.7154739.

[12] Peter Calhoun et al. "Random Forest". In: Nov. 2021, pp. 1–20. ISBN: 9781118445112. DOI: 10.1002/9781118445112.stat08287.

[13] Nitesh Chawla. "C4. 5 and imbalanced data sets: investigating the effect of sampling method, probabilistic estimate, and decision tree structure". In: (Jan. 2003).

[14] Nitesh Chawla et al. "SMOTE: Synthetic Minority Over-sampling Technique". In: *J. Artif. Intell. Res. (JAIR)* 16 (June 2002), pp. 321–357. DOI: 10.1613/jair.953.

[15] Jun Chen and Meng Li. "Chained Predictions of Flight Delay Using Machine Learning". In: (Jan. 2019). DOI: 10.2514/6.2019-1661.

[16] Sun Choi et al. "Prediction of weather-induced airline delays based on machine learning algorithms". In: (Sept. 2016), pp. 1–6. DOI: 10.1109/DASC.2016.7777956.

[17] Jens Clausen et al. "Disruption management in the airline industry—Concepts, models and methods". In: *Computers Operations Research* 37 (May 2010), pp. 809–821. DOI: 10.1016/j.cor.2009.03.027.

[18] Leonardo Cruciol et al. "In-flight cost optimization for air traffic flow management using data mining method on big data". In: *OPT-i 2014 - 1st International Conference on Engineering and Applied Sciences Optimization, Proceedings* (Jan. 2014), pp. 1491–1498.

[19] Yi Ding. "Predicting flight delay based on multiple linear regression". In: *IOP Conference Series: Earth and Environmental Science* 81 (Aug. 2017), p. 012198. DOI: 10.1088/1755-1315/81/1/012198.

[20] Vicente García et al. "The class imbalance problem in pattern classification and learning". In: *II Congreso Español de Informática* (Jan. 2007).

[21] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. "Overview of Supervised Learning". In: Apr. 2009, pp. 1–33. ISBN: 978-0-387-84857-0. DOI: 10.1007/b94608_2.

[22] Tin Ho. "The Random Subspace Method for Constructing Decision Forests". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20 (Sept. 1998), pp. 832–844. DOI: 10.1109/34.709601.

[23] Hassan Khaksar and A. Sheikholeslami. "Airline delay prediction by machine learning algorithms". In: *Scientia Iranica* 26 (Dec. 2017). DOI: 10.24200/sci.2017.20020.

[24] A. Klein, C. Craun, and R.S. Lee. "Airport delay prediction using weather-impacted traffic index (WITI) model". In: (Nov. 2010), 2.B.1–1. DOI: 10.1109/DASC.2010.5655493.

[25] Niklas Kohl et al. "Airline disruption management—Perspectives, experiences and outlook". In: *Journal of Air Transport Management* 13 (May 2007), pp. 149–162. DOI: 10.1016/j.jairtraman.2007.01.001.

[26] Ovidiu Listes and Rommert Dekker. "A Scenario Aggregation Based Approach for Determining a Robust Airline Fleet Composition". In: *Transportation Science* 39 (June 2002). DOI: 10.1287/trsc.1040.0097.

[27] Eric Mueller and Gano Chatterji. "Analysis of Aircraft Arrival and Departure Delay Characteristics". In: (Oct. 2002). DOI: 10.2514/6.2002-5866.

[28] Mykola Pechenizkiy. "The Impact of Feature Extraction on the Performance of a Classifier: kNN, Naïve Bayes and C4.5". In: (May 2005), pp. 268–279. DOI: 10.1007/11424918_28.

[29] Juan Rebollo and Hamsa Balakrishnan. "Characterization and prediction of air traffic delays". In: *Transportation Research Part C: Emerging Technologies* 44 (July 2014), pp. 231–241. DOI: 10.1016/j.trc.2014.04.007.

[30] Hanif Sherali, Ebru Bish, and Xiaomei Zhu. "Airline fleet assignment concepts, models, and algorithms". In: *European Journal of Operational Research* 172 (July 2006), pp. 1–30. DOI: 10.1016/j.ejor.2005.01.056.

[31] Pavel Stefanovič, Olga Kurasova, and Rokas Štrimaitis. "Prediction of Flight Time Deviation for Lithuanian Airports Using Supervised Machine Learning Model". In: *Computational Intelligence and Neuroscience* 2020 (Oct. 2020). DOI: 10.1155/2020/8878681.

[32] Alice Sternberg et al. "A Review on Flight Delay Prediction". In: (Mar. 2017).

[33] Balasubramanian Thiagarajan et al. "A machine learning approach for prediction of on-time performance of flights". In: (Sept. 2017), pp. 1–6. DOI: 10.1109/DASC.2017.8102138.

[34] Yufeng Tu, Michael Ball, and Wolfgang Jank. "Estimating Flight Departure Delay DistributionsA Statistical Approach With Long-Term Trend and Short-Term Pattern". In: *Journal of the American Statistical Association* 103 (Feb. 2008), pp. 112–125. DOI: 10.2139/ssrn.923628.

[35] Varsha Venkatesh et al. "Iterative machine and deep learning approach for aviation delay prediction". In: (Oct. 2017), pp. 562–567. DOI: 10.1109/UPCON.2017.8251111.

[36] T.W.M. Vossen, R. Hoffman, and A. Mukherjee. "Quantitative Problem Solving Methods in the Airline Industry". In: *International Series in Operations Research Management Science* 169 (Jan. 2012), pp. 385–453.