Predicting Flight Delays: How Weather and Seasons Affect Air Travel?

MSc Research Project
Data Analytics

Aniket Guru
Student ID: x22119914
x22119914@student.ncirl.ie

School of Computing
National College of Ireland

Supervisor: Bhart Agrawal

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Aniket Guru
Student ID: x22119914@student.ncirl.ie	x22119914
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Bhart Agrawal
Submission Due Date:	14/12/2023
Project Title:	Predicting Flight Delays: How Weather and Seasons Affect Air Travel?
Word Count:	6588
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	14th December 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Predicting Flight Delays: How Weather and Seasons Affect Air Travel?

Aniket Guru
x22119914
x22119914@student.ncirl.ie

Abstract

This study presents a comparative analysis of LSTM and Random Forest classifiers in predicting flight delays at JFK International Airport, incorporating weather and holiday data for enhanced prediction accuracy. The research demonstrates the efficacy of a 5-class classification system over traditional regression models. The classification approach effectively handles imbalanced data and outliers, offering interpretable insights into delay causatives. The Random Forest model slightly outperforms LSTM with an overall accuracy of 0.9239 in case of unsampled data , while LSTM excels in precision and F1-scores for certain delay classes. Key findings underscore the significant impact of weather, particularly precipitation and wind gusts, on flight delays, with notable disruptions during peak travel seasons. These results equip airlines and aviation authorities with actionable insights to mitigate delays, optimize operations, and improve passenger experience. The novel inclusion of holiday data alongside weather variables in the predictive models contributes to the broader understanding of flight delay factors, paving the way for more informed decision-making in the aviation industry.

Keywords:- LSTM Classifier, Random Forest Classifier,Accuracy , F1-Score

1 Introduction

1.1 Background

Indeed, airborne transport in the twenty-first century is generally quicker, easier, and safer.According to the report published by Bureau of Transportation Statistics (2022)indicates that there was an increase of 21 percent in American airline passengers from 2021 to 2022 as more than 194 million additional persons flew in 2022 . The return of normalcy after the COVID period might have brought about this substantial increase.According to the report publish by Federal Aviation Administration (2023) states year-over-year growth in international traffic at roughly 2.2 percent and approximately four percent in domestic traffic.The industry must remain vigilant and proactive, with providing a better customer experience being key in that respect. One of the main headaches in this industry continues to be flight delays.

Flight delays are now an everyday phenomenon in modern airline transportation. These are frustrating occurrences that delay passengers' journeys, yet they have not been adequately addressed in past research on airline management. Flight delays constitute a serious problem for airlines and passengers hence, it is important for airlines and passengers to understand what causes delays in flights. These delays can be predicted and analysed, and precautions steps can be taken so as to diminish their effect and enhance on-air trips.

1.2 Motivation

Tackling the challenge of flight delays and creating better forecasting tools for advanced scheduling is essential in the aviation sector. Improvements in this area can greatly improve the level of service offered. This in turn, could lead to happier customers and which will lead to boost the operational effectiveness and financial success of airlines.
According to Ball et al. (2010) Flight delays are a big problem in the U.S., if we talk about Back in 2007, about one out of every four flights got to their destination over 15 minutes late . A lot of these delays happened because the aviation system couldn't handle all the traffic. Another significant portion of the delays stemmed from issues within the airlines themselves. The majority of other delays occurred when a plane arrived late, subsequently causing its subsequent flight to depart behind schedule.
According to the U.S. Department of Tourism it is about a hundred and ten dollars per minute to operate an airline. As a result, every minute a flight is delayed, each of this amounts to additional dollars spent by the airline. For example, a 15-minute-late flight would cost about $1,650 while barely being considered late. According to Bureau of Transportation Statistics (2022) released by the three busiest U.S. airports , twenty-one percent of flights from said airports that went to different destinations in America took more than fifteen minutes to arrive. On the other hand, Department of Transportation sees "on time" as anything less than 15 minutes delay on landing..

1.3 Importance of John F. Kennedy International (JFK)Airport.

Located in NYC, JFK is the busiest airport in the city and has 5 passenger terminals. In all, over 70 airlines have routes in and out of JFK, and it's a base for heavy hitters like Delta, American Airlines, and JetBlue. JFK Airport has 4 runways and covers 5,200 acres in Queens, and all this space is definitely required to shuttle the over 55 million annual passengers it welcomed in 2022 alone.According to article published by Lancelotte (2023) New York's JFK Airport had the highest CDF score of 61.58%, and it also has the longest LDA90 (Long Delayed Arrival By 90+ Minutes)at 104 minutes}.

1.3.1 Operational Challenges at JFK

JFK faces a range of operational hurdles due to factors like seasonal shifts, diverse weather conditions, and fluctuating patterns of holiday travel. This project aims to delve into and forecast flight delays at JFK International Airport, taking these critical aspects into account. At JFK, seasonal variations significantly impact flight scheduling and delays. Different seasons, including summer, winter, autumn, and the rainy season, bring unique operational challenges. In the summer, lots of people travel, so flights often get delayed with all the extra passengers and planes. In the winter, things like snowstorms and really cold weather can cause a bunch of delays too.

1.3.2 Environmental Challenges at JFK

Weather conditions are a pivotal factor affecting flight schedules at JFK. The airport is in a spot where bad weather like thunderstorms, fog, and snowstorms happens a lot. This can really mess up flights. For example, the area's frequent storms can cause flights to be delayed or even cancelled to keep everyone safe.. Foggy conditions, reducing visibility, also adversely affect landing and take-off operations. By analysing historical weather data and its relationship with flight delays, insights can be drawn on how different weather scenarios affect the airport's functionality.

Holiday periods also influence flight delays at JFK. High-travel seasons such as Thanksgiving, Christmas, and New Year's lead to airport congestion, stretching the airport's capacities. The spike in passenger flow during these times can strain airport resources and contribute to delays in various processes, including check-in, security, and baggage handling. Examining flight data from past holidays provides an understanding of patterns contributing to these delays, aiding in better preparation and management.

1.4 Research Aim

To conduct a thorough investigation, we have amassed flight data covering a 7-year period, from 2017 to 2022. This expansive data-set allows us to scrutinize current trends, pinpoint recurrent patterns, and assess the success of previously implemented strategies aimed at reducing flight delays. By analyzing historical data, we can construct predictive models that consider time-related variations and accurately forecast future flight delays. In this study, our primary research objective is to delve into the complex relationship between flight delays at John F. Kennedy International Airport (JFK) and various determinants, including seasonal changes, meteorological conditions, and holiday travel behaviour's, using a 7-year dataset from 2017 to 2022. Our goal is to develop predictive models employing a 5-class classification system to efficiently foresee flight delays at JFK and to gain an in-depth understanding of the influence of different weather and seasonal factors on flight delays. Furthermore, we will tackle the issue of unbalanced data and endeavour to enable practical applications in the aviation sector, ultimately aiming to boost operational effectiveness, reduce disruptions, and enhance the overall travel experience for passengers at JFK.

2 Related Work

Flight delays are considered to be one of the most important indicators of performance in the aviation industry's air transportation systems. Unexpected alterations in meteorological circumstances frequently hinder the seamless operation of air transportation systems.

Kim et al. (2016), came up with an new way to figure it out. If flights will be late because of the weather or not. They used the flight data from inside the U.S. from 2005-2015 that they got from the Bureau of Transportation Statistics. They mostly looked at flights landing in Charlotte Douglas International Airport from Denver International Airport. They used a bunch of learning methods that use past data to make predictions, and mixed up flight times with weather stuff at airports. Things like what part of the month it is, the day, when the plane leaves and arrives, and if it's usually late were important.

They also looked at the weather like wind direction and speed, how much you can see, rain, snow, and more. They picked methods like decision trees, random forest, AdaBoost, and k-nearest neighbors to make their guessing model. To see if they were right, they did 10-fold cross-validation and used ROC curves.

Zhang and Ma (2020), worked on a different model using BTS data. They made a model to guess flight delays when leaving Newark Liberty International Airport from January 2016 to December 2017. They got weather info from Local Climatological Data (LCD).Similarly to Kim et al. (2016), they also utilize features like when the flight leaves, delay times, the airline code, month, quarter, day of the week, how far the flight is, and weather conditions, rain, how well you can see, wind speed, air pressure, and ceiling. However, they have used a different method, called CatBoost, for their model creation. They got it right about 77% of the time.

The study by Borsky and Unterberger (2019) analyzed flight departure data from ten major U.S. airports, spanning from January 2012 to September 2017. The research focused on developing a model to predict departure delays. This approach differed from previous studies by categorizing weather conditions into two types: sudden onset events (such as precipitation and wind) and slow onset events (like temperature changes). The team applied a Prais-Winstein estimator with panel-corrected standard errors to assess the impact of these weather events on flight delays. Their findings indicated that weather disturbances could lead to an increase in departure delays by up to 23 minutes.

The study by Addu et al. (2022) (2022) examined historical flight and weather data in the U.S. over a five-year period. Their flight dataset included details like departure and scheduled times, arrival time and delay, flight number, airports of origin and destination, and the carrier. Weather data parameters included temperature, wind direction and speed, pressure, and the time and hour. Adopting a similar approach to earlier research, they employed multiple algorithms such as Random Forest, Decision Trees, and KNN to develop their prediction model. Additionally, they used Naïve Bayes and MLP Classifier, specifically focusing on predicting departure delays. Distinct from previous studies, they created a webpage to display statistics and predictions for both arrival and departure delays. Identically to previous studies , Jiang et al. (2020)also applied machine learning techniques to predict flight delays within the U.S. airspace for the year 2016. They utilized data such as scheduled and actual departure and arrival times, carrier information, origin and destination airports, airtime, and non-stop distance, sourced from the BTS. Their weather data was obtained from the Quality Controlled Local Climatological Data (QCLCD), focusing on visibility and wind speed as key features. Setting their work apart from studies like above research, their most successful model, which achieved a prediction accuracy of 89.07%, was developed using a Multilayer Perceptron.

Mokhtarimousavi and Mehrabi (2023)employed a statistical method to analyze the primary causes of flight delays at Miami International Airport during the peak delay months of June, July, and August in 2019. They utilized the BTS dataset, incorporating factors such as departure delay, taxi out time, NAS weather, origin airport, day of the week, month, weather, and carrier delay. In contrast to studies, their approach involved using a Support Vector Machine (SVM) model, which was trained using an Artificial Bee Colony (ABC) algorithm. This method was applied to investigate the non-linear connections between various factors and the outcomes of flight delays.

Yu et al. (2019) analyzed comprehensive data from the Beijing International Airport, covering the data from January 2017 to March 2018, to identify key factors contributing to flight delays and to develop a flight delay prediction model. Their study included standard airport data features such as airline information, air traffic control details, aircraft types, check-in and flight closing times, boarding and take-off times, as used in previous studies listed above. However, they uniquely incorporated less commonly considered factors like actual airport crowdedness, air route conditions, aircraft capacity, and various boarding methods. Study employed an innovative deep belief network approach to uncover underlying patterns in flight delays, integrating Support Vector Regression for supervised fine-tuning in their model. Their results showed that 99.3% of the predicted delays were within a 25-minute margin of error compared to actual observed values.

Li and Jing (2022).focused on analyzing flight data from domestic flights in China during the months of June to August 2016, aiming to predict flight delays by examining both spatial and temporal aspects. They first constructed an aviation network using graphs, nodes, and edges to identify spatial characteristics. Then, following a similar approach to Yu et al. (2019),they incorporated temporal factors such as airport and aviation crowdedness, along with past and future weather conditions. However, in contrast to above studies .They employed a Long Short-Term Memory (LSTM) method combined with a random forest algorithm to forecast flight delays.

Gui et al. (2020),adopted a distinct method for predicting flight delays. Their strategy involved constructing a dataset from automatic dependent surveillance-broadcast (ADS-B) messages. These messages were received, pre-processed, and then combined with additional information like weather conditions, flight schedules, and airport details. Like studies [7] andLi and Jing (2022),they utilized flight and weather data sourced from the Civil Aviation Administration of China (CAAC). In terms of methodology, they followed a similar approach as above study,employing algorithms like Long Short-Term Memory (LSTM) and Random Forest in their predictive model.

Ye et al. (2020),developed a novel approach for forecasting overall flight departure delays at airports, employing supervised learning techniques. Their methodology was akin to Gui et al. (2020),utilizing flight data from the Civil Aviation Administration of China, specifically focusing on data from March 2017 to February 2018 for Nanjing Lukou International Airport. They incorporated publicly available weather data, encompassing elements like dew point temperature, humidity, wind speed, wind direction, and pressure. In addition to utilizing Support Vector Machine (SVM) and Random Forest algorithms, similar to studies previous also applied a Light GBM model for their predictions, achieving an accuracy rate of 86.55%.

Yiu et al. (2021),analyzed flight data from Hong Kong International Airport, specifically from March 31, 2018, to April 30, 2018. They explored various machine learning methods to forecast flight delays. Echoing the approaches . they utilized the Random Forest algorithm but also evaluated their model using other techniques such as K-Nearest Neighbors (KNN), Naïve Bayes, Artificial Neural Networks (ANN), and decision trees. They found that rainfall had the most significant influence on flight delays, with other factors like humidity and air pressure also being key contributors. Among their models, the ANN yielded the highest accuracy, reaching 83.17%.

Borse et al. (2020),considered various weather-related factors like temperature, humidity, rainfall in millimeters, visibility, and month number, along with flight data parameters such as month, day, day of the week, flight number, origin and destination airports, scheduled departure, departure delay, taxi-out time, distance, and scheduled arrival, to develop their flight prediction algorithm. Diverging from other studies, they employed logistic regression, decision trees, and neural network models. Their findings showed that the decision tree model was more effective in predicting on-time flights, while the neural

network was better at forecasting flight delays. Additionally, they created a user interface where users can input the departure city in a text field to view the prediction results.
Kaewunruen et al. (2021),focused their study on flight and weather data from Birmingham Airport and the Meteorological Office, spanning from January 1, 2018, to December 31, 2019. Setting their work apart from studies ,they applied a combination of linear regression, random forest, artificial neural networks, and an SVM model to identify weather factors influencing flight punctuality. They defined punctuality rate as the dependent variable and considered factors like the number of scheduled flights, temperature, wind power, meteorological conditions, month, day of the week, and day of the month as independent variables. Additionally, they utilized grid search and cross-validation techniques to optimize their model parameters. Upon evaluating the models based on RMSE and R^2 values, they found that the Random Forest model was the most effective in predicting the punctuality rate of flights at Birmingham Airport.
Liu et al. (2022),presents a multi classification flight delay prediction model based on long short term memory networks (LSTM) is constructed in this paper. Model combines the flight data and weather data ,make use of the timeseries characteristics in this pape. the three-layer LSTM network structure is compared with SVM and MLP. The number of hidden layer nodes is 128, and the learning rate is 0.01. Batch_ Size is set to 128, epoch is set to 6, and loss function is cross entropy. LSTM outperformed with 96.01 percentage of accuracy.
Research Contribution:-: In this research project , inspired by the four-class classification approach to flight delay predictions reported in Gui et al. (2020) and the five-class classification studies by Liu et al. (2022), Random Forest and LSTM models are being employed on the JFK flight data-set combined with weather and holiday data. This approach aims to enhance the efficiency of the system and create a robust classification model. The inclusion of weather and holiday factors, which were not considered in the aforementioned studies, is pivotal for predicting flight delays accurately.

The decision to adopt a 5-class classification instead of regression for flight delay prediction is based on several key advantages:

- It provides clearer insights into the impact of various factors on delays, thereby improving model interpretability.
- Classification more effectively manages imbalanced data by focusing on each class independently.
- While regression may falter with extreme delay values, classification maintains robustness against outliers.
- This method is well-suited to practical applications, potentially aiding operational improvements in the aviation industry.

3 Methodology

This chapter elaborates on the methodical strategy implemented in this study, delineating the techniques, instruments, and design workflow employed. The investigation adheres to the Cross-Industry Standard Process for Data Mining (CRISP-DM) as its guiding framework.That Includes 6 steps.Fig 1 shows the all steps involved in the process.

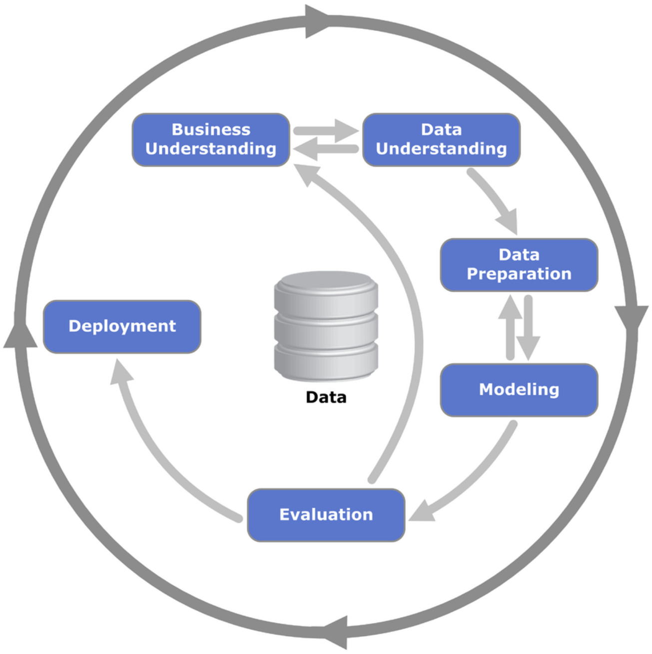

Figure 1: CRISP-DM Process Diagram

3.1 Business Understanding:

At the first phase of the research is dedicated to considering American's airline industry. This involves highlighting significant parts of business such as flight schedules, seasonal variations and determining reasons for delayed flights as well as elements that affect it.

3.2 Data Understanding

My research uses the latest flight data from 2017 to2022 sourced from the U.S. Bureau of Transportation Statistics to ensure a detailed and current analysis. This was done to ensure that the data used to enhance the model's prediction accuracy could be trusted. Besides that, I used Kaggle's US holiday data to measure their impact on the flight schedules and traveling patterns.

3.3 Data Preparation

During this crucial stage, the Python programming language is used in a Jupyter Notebook setting. Data loading and replacement of missing values through the average of the attributes' is done using the Pandas' library. Pattern recognition is increased by data transformation (e.g., integer to float conversion) and visualizations (Matplotlib,Seaborn and Plotly). This phase includes encoding categorical data as numerical values.

3.4 Data Modelling

Subsets represented appropriate ratios for training, testing and validation dataset. By making use of cross-validation techniques, validity of the predictive models was enhanced. Sophisticated methods like LSTMs within the TensorFlow environment and Random Forests implemented by Scikit-learn were employed for the analysis. Every algorithm went through stringent hyper parameter tuning with a purpose of enhancing the prediction accuracy.

3.5 Evaluation

The evaluation of the models, specifically Random Forest and LSTM, in predicting flight delays, was meticulously conducted across the training, validation, and testing datasets. This comprehensive assessment employed a range of metrics, including precision, recall, F1-score, accuracy, and ROC, to gauge the effectiveness of each model in accurately forecasting flight delays. These metrics were chosen for their ability to provide a holistic view of the models' performance, ensuring a robust and well-rounded evaluation.

3.6 Deployment

After thorough evaluation and selection of the optimal predictive method, the research culminates in the deployment stage. The developed solution is ready for use by airlines and passengers to anticipate flight schedules and potential delays.

4 Design Specification

4.1 Proposed LSTM Based Method Structure

There are multiple reasons for flight delay, such as weather, human and other reasons, and a plane may perform multiple flight tasks at the same time, so the punctuality of the previous flight also has a great relationship with whether the flight is delayed. So the dataset of flight delays has a time series. LSTM is a specific cyclic network (RNN) that can capture long-term correlation. LSTM model stands out for its advanced architecture, designed to effectively counter the gradient vanishing challenge often encountered in standard RNNs.
Our LSTM model operates under a core that processes inputs arranged as $x = [x_1, x_2, \ldots, x_T]$. In this case, x_t denotes a specific feature vector at every instant t. At each time step, the model computes a hidden state $h = [h_1, h_2, \ldots, h_T]$, accumulating the information of the sequence present in the whole sequence until t. The output sequence is $y = [y_1, y_2, \ldots, y_T]$, that is, the predictive or transformations output of the model.
Fig.2 shows that the single LSTM cell contains four gate structures, namely forget gate, candidate gate, input gate and output gate.In our Lstm.

- **Forget Gate (`ft`):** This assesses the relevance of previous information in the cellular system, determining that which can be discarded from the cell state thus aiding the system to forget irrelevant information.
- **Input Gate (`it`) and Candidate Gate (`c~t`):** The combined effects of these gates will decide on the new information that shall update the already held values in the cell state.
- **Output Gate (`ot`):** Shaping the next hidden state `ht` that represents previous information as input and on which other outputs depend.

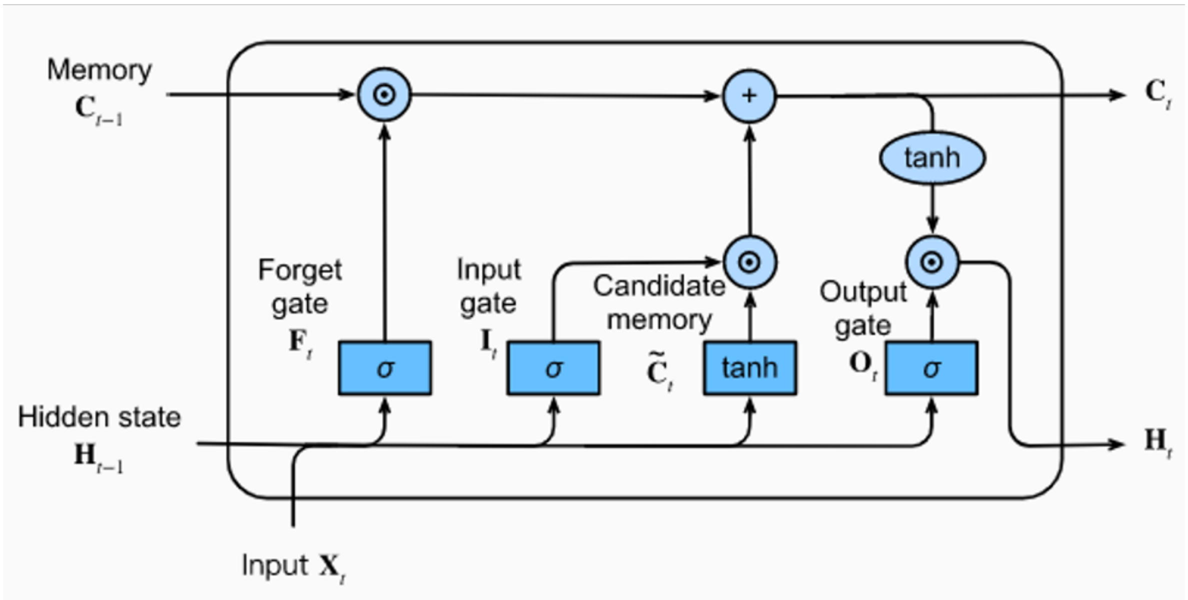

Figure 2: LSTM Based Method Structure

The LSTM model was fine-tuned using grid search with a specific set of parameters is : Units of 50-100, dropout of 0.2, batch size at 32, epoch equals 10. Used in this process, the best possible combination came out after applying hyperparameter tunning as a 100 unit model fed into two LSTMs, with a 0.2 dropout, batched at 32 and trained for 10 epochs.

Regarding the architecture, the model starts with the 100-unit LSTM layer which processes the entire input sequence and provide sequences to subsequent layers by ensuring smooth flow of the data. For a solution to overfitting, which is one of the major challenges in deep learning, a dropout layer at a rate of 0.2 comes out the next. For the classification task that follows, the second LSTM layer of 100 units does not output sequences but concentrates on summation of the input sequence. The second dropout layer is introduced, which repeats the initial dropout layer at the same rate makes the model more robust against generalization. The final stage comprises a dense output layer that has five units featuring softmax activation function for multi-class classification and classification of outputs into one out of five categories.

4.2 Proposed Random Forest-Based Model

The Random Forest model excel in multi-class classification for the tasks like predicting flight delays due to its robustness to over-fitting and its ability to handling the heterogeneous data. It can capture complex feature like weather data and flight delay data and perform the interactions without extensive prepossessing, making it ideal for the non-linear and multi-factorial nature of flight delay cause. Additionally, it is inherent feature importance evaluation provides valuable insight for understanding the key features causing the delays in flight departure.

The pre-trained Random Forest model underwent fine-tuning through grid search cross-validation, applying a targeted parameter grid with options for the number of trees (`'n_estimators': [100, 200]`) and tree depth (`'max_depth': [None, 10, 20]`). This hyperparameter optimization process yielded the optimal model configuration as {`'max_depth': None, 'n_estimators': 200`}, indicating an unrestricted tree depth and a comprehensive ensemble of 200 trees, which is optimal for capturing the complex patterns in the dataset without overfitting.

5 Implementation

In this section , the paper discussed the implementation of the research starting from the data collection to model building. Fig 3 represents the block diagram of the workflow followed.

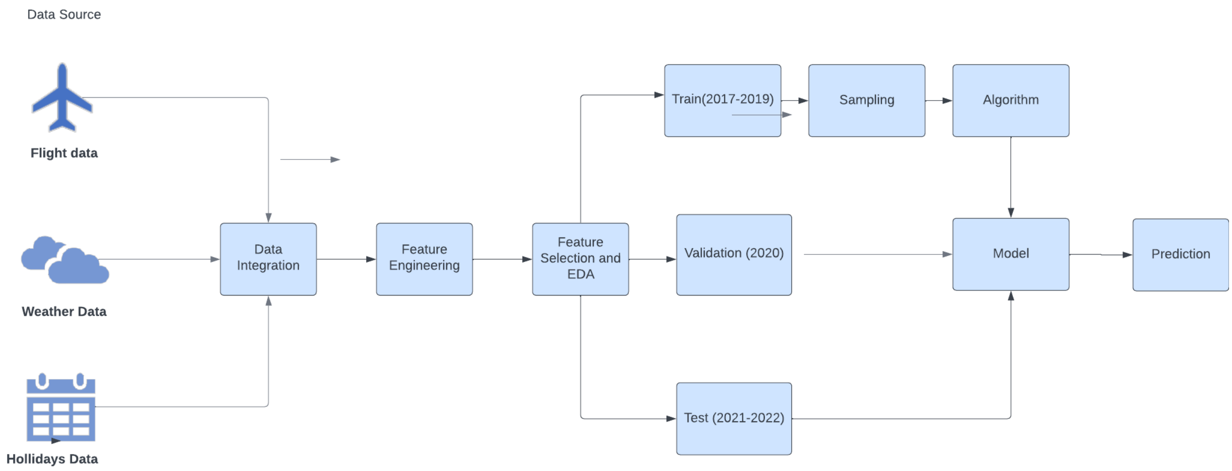

Figure 3: Flight Delay Prediction Workflow: From Data Collection to Model Deployment

5.1 Data Selection

In this project,uses flight data spanning from 2017 to 2022 to maintain the pertinence and precision of our analysis. The recent data has been selected to aim to get better results. The study uses three different data sets sourced from opensource data sources.
The data regarding flights was procured from the Bureau of Transportation Statistics (2023), which is recognized as the central repository for U.S. transportation data. This dataset is comprehensive, including details such as departure times, flight identification, destinations, both planned and actual departure timings, durations of flights, and a variety of delay types, to list some. Overall, the dataset encompasses close to 4,00,000 entries.
For weather-related insights, we turned to the Open-Meteo (2023) , a resource for historical weather recordings taken every hour, available from as far back as 1940. From this extensive collection, we selected weather variables that align with the goals of our study, specifically temperature, wind metrics, cloudiness, and precipitation.
U.S. holidays were sourced from Kaggle (2023), which provides an exhaustive overview of celebrated holidays in the U.S. It details when each holiday occurs, what it is called, and which day of the week it is celebrated. This information is pivotal for our investigation into how holidays affect flight schedules and to delve deeper into the trends and patterns associated with holidays.This data set contains the list of holidays from year 1722 to 2122. But for our uses filtered the data from 2017 to 2022.

5.2 Data preparation

This stage encompasses the preliminary actions taken to ready the dataset for subsequent processing.

5.2.1 Integrated Flight, Holliday and Weather Data System.

Our study was built on the foundation of merging flight and weather data. To this end, we have noted weather's daily readings which we matched with flight data according to their respective dates and timings. We went ahead with various time settings including UTC and EST to enhance accurate data fusion. It also availed us an opportunity of exploring how it can impact on delayed flights owing to bad weather conditions.
We also added the Holiday component and used it in conjunction with departure date which was associated with particular holiday dates. In our analysis, we primarily considered the day's name and date of the holiday.

5.2.2 Refining the Data

During the data preparation phase, it turned out that in the case with the US Bureau of Transportation Statistics, the majority of the available data could be used directly without additional processing. We practiced data cleaning practices including converting dates into the accepted forms as well making everything lowercase. To replace holidays data for the missing values we used "no holidays." Additionally, we added a correction factor to flight delay data whereby we changed negative values denoting early-departure flights to zero. We verified our data to achieve consistency so as to use it in other subsequent studies and modelling.

5.3 EXPLORATORY DATA ANALYSIS

In this part of the research, Exploratory Data Analysis was conducted in order to plot the data, find a trend if it is identifiable and examine relationships between different variables. The findings from this exercise shall form the fundamental basis for the other subsequent analytical exercises mentioned in our research study

5.3.1 Distribution Of Delay Classes

The distribution of various delay categories within each class of delay was depicted through a bar chart in fig 4. The data reveals a notable imbalance among classes: the "no delay" category comprises the majority with over 370,000 flights, in stark contrast to the "Extreme" delay class, which includes around 3600 flights.

5.3.2 Seasonal effects on departure delays

The stacked bar shown in Fig 5 monthly averages of delays caused by weather and total average departure delay in minutes of yearly data on departure delay caused by weather and other circumstances at the JFK Airport. It is structured as a stacked bar chart, where two types of delays are shown: weather-related delays and departure delay for a single flight.
The chart in Fig 5 reveals that June and July were those months with greatest total departure delays and additionally significant weather-type delays for the month of July.

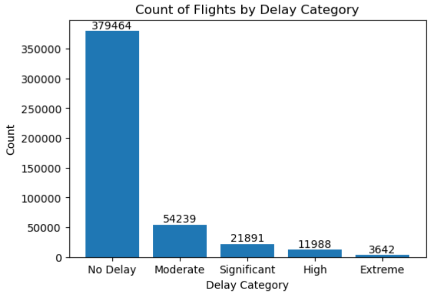

Figure 4: Delay Category

Thus, summers with famous thunderstorms and sometimes violent weather disrupt more flights in JFK Airport than any other time of a year. On the contrary, the month of April, September, October, November, December have less average delays caused by weather, hence probably more favourable weather conditions for airborne flights. These are indicators of how real weather plays a role in flight punctuality since some seasons and seasons are more associated with weather-related delay.

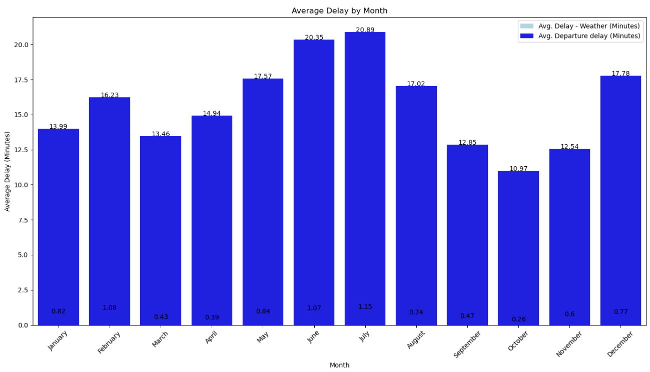

Figure 5: Average Delay by Month

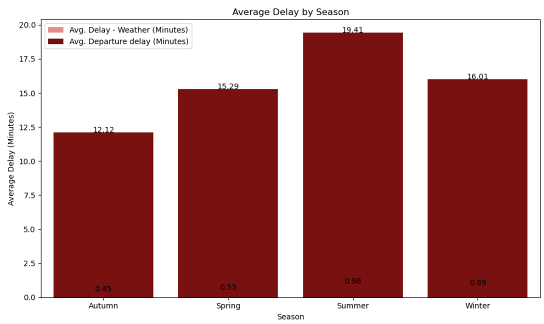

Figure 6: Average Delay by Season

The Fig 6 demonstrates seasonal flight delays at JFK Airport that clearly show that the inclement weather conditions in autumn play a relatively insignificant part in causing the delays, whereas summer and winter weather has substantial impact with delays caused by thunder storms in summer and snow and Softer conditions are observed during spring and autumn compared to summer which causes lesser weather-related delays

5.3.3 Holidays and Weekend effect on flight delays

This fig 7 shows monthly mean departures delays from New York's JFK airport during holiday and non-holidays every month. Holidays also contribute greatly to the delays because of the New Years and Christmas trips. It is observed that the highest increases are registered for January and December with respect to these situations. Holiday- related travel will probably result in significant variations in flight schedules during this seasonal trend owing to high passenger loads, as well as possible unpleasant winter climate.

Fig 8 shows weekends in January and December at JFK experience more severe holiday congestion as well as the weather-related snowfalls and ice storms, thereby causing bigger traffic backups than during weekdays. More travel disturbances expected during the holiday season.

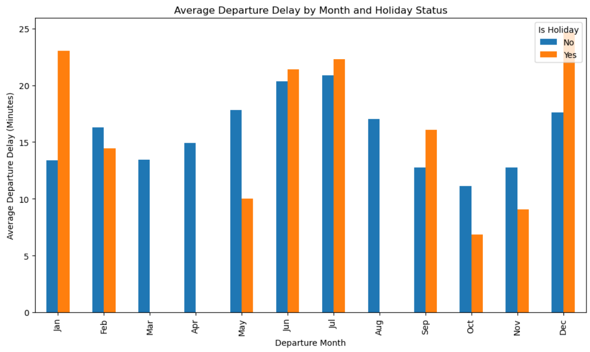

Figure 7: Average Departure Delay by Month and Holiday Status'

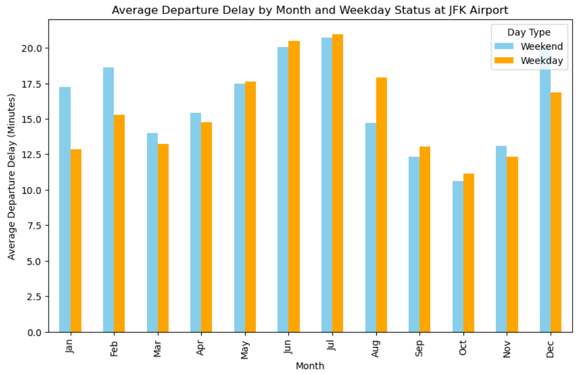

Figure 8: Average Departure Delay by Month and Weekday Status at JFK Airport

5.3.4 Effects of Weather on Departure Delay in 2022

The Fig 9shows the daily correlation of delay flights delays avg and precipitation avg as well as gust avg in 2022. Nevertheless, summer has nothing to do with delayed signal since there is less precipitation and strong wind, low turbulence; days numbering from 180-280. Summer weather also suggests that there are clues of irregular punctuality in JFK's flight. On the other hand, summer delays cannot be associated with the amount of rainfall that falls in a month and speed of wind gusts; hence the only possible explanation could include factors like wind speed and clouds coverage or even temperature influences. This therefore implies that precipitation and wind gusts are likely to be among the critical variables considered in JFK Airport's delay predictive model as they play a significant role when it comes to air travel delays

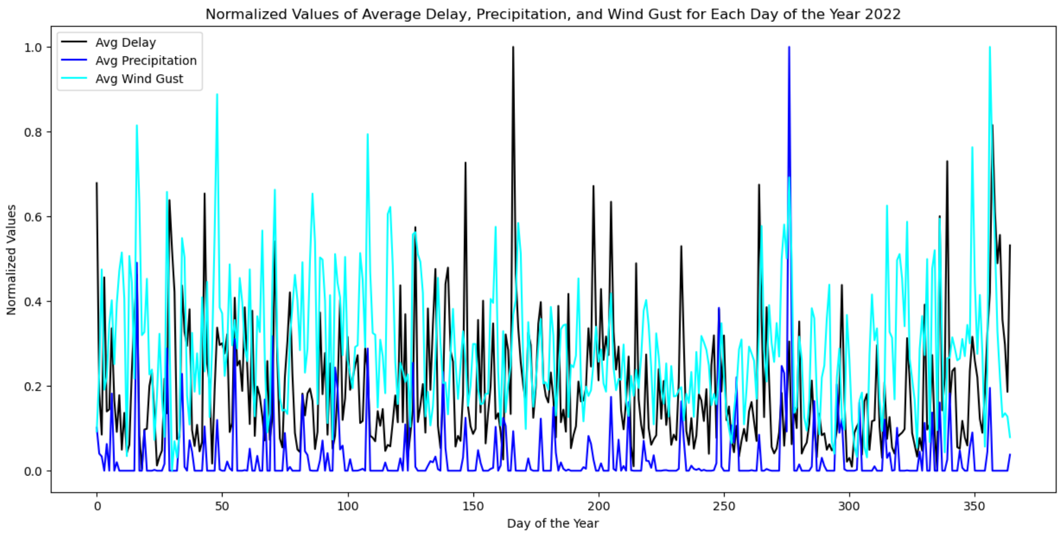

Figure 9: Effects of Weather on Departure Delay in 2022

5.3.5 Correlation

While checking for correlation with the Pearson correlation method, it determines whether continuous related variables are positively or negatively correlated. A strong relationship of 0.8 obtained in this study shows that often increasing wind speeds are accompanied by

increased wind gust intensities.We have also checked the which feature has multicollinearity more than 5 and in that Scheduled elapsed time (Minutes) & Actual elapsed time (Minutes) toped with value more than 0.95.
Cramer's V, on the other hand, is based on categorical data and represents the association between two variables, while Cramer's V ranges from zero to one. An almost perfect relationship of 0.97was established between tail number and Carrier code. It may be the case because most airlines assign distinct flight numbers to different aviation sectors. The two tests are vital for the understanding of various correlative relations in the aviation data and they give a clear view of their optimization process.

5.4 DATA PRE-PROCESSING

In this section, After the integration and exploratory data analysis (EDA) of flight, weather, and holiday data, further data processing steps were conducted as follows.

5.4.1 Feature Extraction

As part of feature extraction, several features were derived from the date and time columns to capture important temporal information Table 1. Next, we converted the prediction task into a 5-class classification problem, as inspired by relevant research Jiang et al. (2020).The flight delay(minutes) was categorized into five intervals capturing different levels of delay severity Table 2. This approach provides a more nuanced understanding of flight delays, enhancing the accuracy and effectiveness of our delay prediction model.

Table 1: Features Extracted

Original Feature	Extracted Feature
Departure Delay	Delay Category, Avg Delay Previous Hour
Departure Date	Day, Month, Season, Year, Day of the Week
Departure Time	Departure Hour & Minute, Flights per Hour
Holiday Name	Is Holiday

Table 2: Flight Delay Classification

Flight Delay (mins)	Delay Class Code	Delay Class
$T \leq 15$	0	No Delay
$15 < T \leq 60$	1	Moderate
$60 < T \leq 120$	2	Significant
$120 < T \leq 240$	3	High
$T > 240$	4	Extreme

5.4.2 Feature Selection

The selectkbest method from sckit learn library was used for feature selection select key indicators that can forecast flight delay. Such a rule uses a score function in assigning the weight of the features. it ranks the features according to their significance and then picks its first k features.D with the highest scores. Therefore, we listed down the major fifteen key aspects. the highest scores Table III. These features were considered to have the major effect in forecasting of flights delay.

5.4.3 One-hot Encoding

This research uses one-hot encoding and manual labeling in feature encoding. Those features, 'Departure Days', 'Is Holiday' and 'Delay Class' with a smaller number of unique values were encoded manually. One-Hot encoding technique was used for converting large numbers of unique values from categorical to numeric format. Using one-hot encoding on other categorical variables made it possible for the model to handle categorical data more efficiently.

5.5 MODELLING

The modelling section describes the approaches and techniques employed to address the research and one of the important steps and the features which are being used in the models are listed in table. models, and evaluate their performance.. Table 3 shows the input variable to model.

Table 3: Model Input Variables

Independent Variables	**Dependent Variable**
Scheduled elapsed time (Minutes),Taxi-Out time (Minutes)	Delay Class
Delay Carrier (Minutes) ,Delay Weather (Minutes)	
Delay National Aviation System (Minutes) , Delay Security (Minutes)	
Delay Late Aircraft Arrival (Minutes) ,Departure Month	
Departure Day ,Departure Hour,Weekday	
Weather code (wmo code) ,Cloud cover (%), precipitation (mm)	
Wind speed 100m (km/h) ,Wind direction 100m (°)	
Wind gusts 10m (km/h) , Season	
Avg Delay Previous Hour ,Flights per Hour	
Destination Airport ,IS_Holiday	

5.5.1 Data Splitting

The dataset was divided into three Categories , namely train, validate, and test, with careful consideration of temporal aspects like the size of training testing and validation. The training set was constructed using historical data spanning from 2017 to 2019, while the validation set comprised data from the year 2020. Furthermore, the test set consisted of data from the year recent years like 2021 and 2022, ensuring that the model's performance was assessed on future data. By adopting this temporal division, we were able to evaluate the model's generalizability and its ability to effectively handle unseen future data in real-world scenarios.

5.5.2 Sampling

As the EDA Suggested that there is a high degree of class imbalance in the dataset. To address this issue, we used techniques to re-balance the distribution between majority and 4 minority classes, as correctly identifying instances from the minority class is crucial in real-world scenario. From the Various over and under sampling techniques SMOTE, is being employed to balance the classes like no delay, moderate, significant, high, and extreme delays respectively. addressed the imbalanced data challenge and improved the overall performance of the model.

6 Evaluation

During assessment of the models, the evaluation was conducted on them. by means of the train, validate, and test data sets. different measures including precision, recall, F1-score, and accuracy. ROC.

6.1 Experiments

6.1.1 Experiment Result Based on LSTM and Random Forest Classifier

This section explain the evaluation matrix of the both models on original data set without doing the oversamplling.
Based on the Table 4 the LSTM and Random Forest classifiers show similar performance with slight differences in the parameter, Both are performing well in all aspects while there are few differences as well . The LSTM classifier boasts an accuracy of 0.92, slightly lower than Random Forest's 0.9239. In class-wise performance, both models excel in different areas: LSTM leads in precision and F1-score for certain classes, while Random Forest shows better recall and F1-score in others. Although both have an equal weighted average F1-score of 0.91, Random Forest has a slight edge in overall accuracy.

Table 4: Classification Report for LSTM and Random Forest

LSTM				Random Forest			
Class	Precision	Recall	F1-Score	Class	Precision	Recall	F1-Score
No Delay	0.94	1.00	0.97	0	0.94	1.00	0.97
Moderate	0.90	0.57	0.70	1	0.88	0.61	0.72
Significant	0.82	0.72	0.77	2	0.84	0.81	0.81
High	0.74	0.91	0.81	3	0.91	0.79	0.80
Extreme	0.88	0.77	0.82	4	0.89	0.74	0.59
Accuracy		0.92		Accuracy		0.9239	

While looking at the confusion matrix in fig 10both classifiers perform exceptionally well for the 'No Delay' class, the Random Forest classifier has a slight edge in accurately predicting the 'Moderate', 'Significant', and 'High' classes with fewer misclassifications across the board. The LSTM model, however, seems to have a slightly better true positive rate for the 'Extreme' class.
However, LSTM's specific class strengths make it a preferable choice depending on the importance of accurately predicting certain classes or the impact of misclassifications in practical applications.

6.1.2 Based on Over sampled data

Both models with their best parameters are being applied to the over-sampled data-set, and evaluation is being done based on their performance on the test data-set.
Based on the Tabel 5 LSTM classifier slightly outperforms the Random Forest classifier, achieving an overall accuracy of 0.92 compared to 0.91. While both models exhibit high

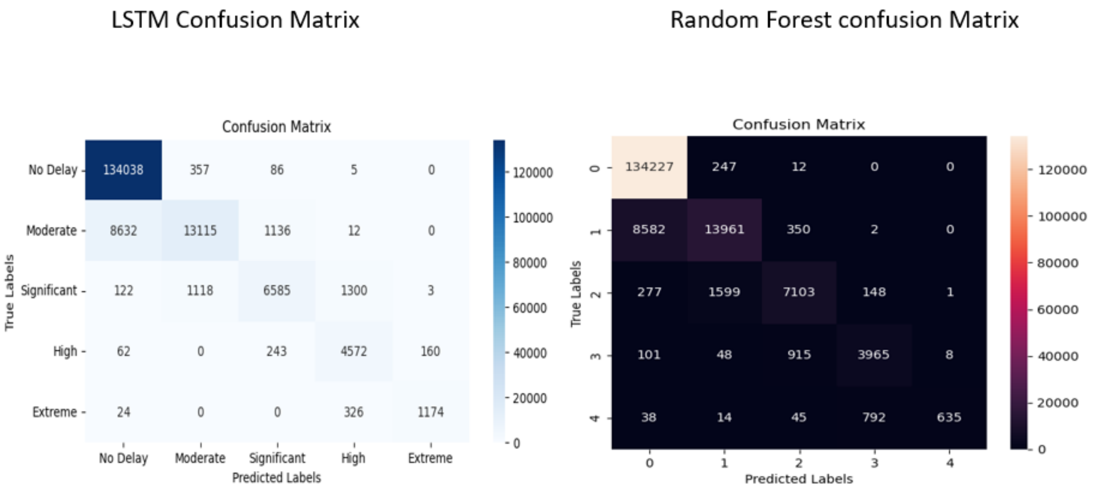

Figure 10: Confusion Matrix

precision and recall in the 'No Delay' category with an F1-Score of 0.96, differences emerge in other classes; notably, LSTM leads with a higher F1-Score in the 'Extreme' delay category (0.91 vs. 0.88). Despite the Random Forest's marginally better macro average of 0.89, indicating consistent performance across all classes, the models share an identical weighted average F1-Score of 0.92, suggesting similar efficacy when accounting for class distribution.

While looking at confusion matrix on fig 11for all class LSTM has better true positive

Table 5: Classification Report for Oversampled LSTM and Random Forest

Class	LSTM Oversampled Classifier			Random Forest Classifier		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
No Delay	0.94	0.98	0.96	0.94	0.98	0.96
Moderate	0.83	0.58	0.68	0.79	0.58	0.67
Significant	0.80	0.89	0.84	0.81	0.84	0.85
High	0.91	0.85	0.88	0.87	0.81	0.88
Extreme	0.89	0.93	0.91	0.94	0.83	0.88
Accuracy	0.92			0.91		

rates than random forest. Both are performing well in classifying "No delay" class. especially in the 'Extreme' delay category. The RF classifier, although showing robust numbers, indicates a slight difficulty with the 'Extreme' delay category, as seen by the lower true positive rate.

However, the LSTM classifier is very good in predicting "No Delay" and "Moderate" thus hints that it might be excellent with easy delays. In contrast to this, however, Random Forest classifier reveals a relatively consistent performance in all cases with exceptional attention into 'Significant' and 'Extreme' delays. This means that it could be the better option of predicting the more extreme types of delay.

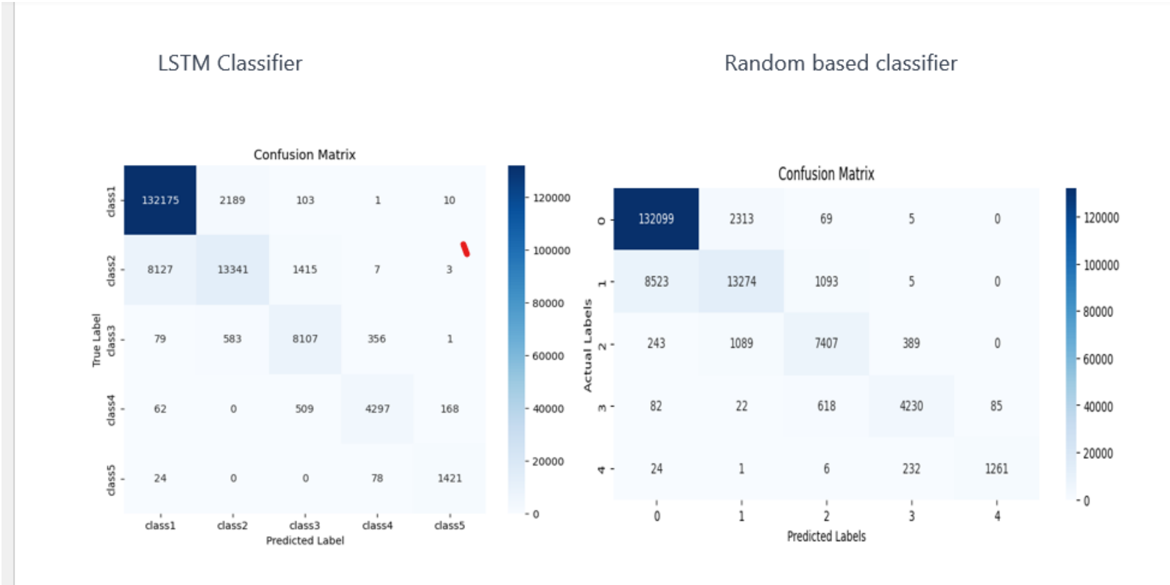

Figure 11: Confusion Matrix

6.2 Discussion

Results from our LSTM and RF classifiers surpass those from the study Gui et al. (2020), achieving 92.2% accuracy in multi-class classification compared to their 90.2% in binary and below 70% in four-class settings. Additionally, our research utilized a more comprehensive 5-class categorization for flight delays, an enhancement over their 4-class system. Furthermore, our models outperformed the study Jiang et al. (2020), which reached only 80% accuracy with various machine learning models, including a multilayer perceptron at 89%. This demonstrates our LSTM and RF models' significant advancement in delay prediction modeling.
The RF classifier demonstrates superior performance in 'Significant' and 'Moderate' delay categories with unsampled data, while LSTM excels in 'Extreme' and 'High' delays, particularly with oversampled data. This contrast highlights the necessity for continuous improvements in feature extraction and RF model development, aiming to refine flight delay classification for more effective air traffic management.

7 Conclusion and Future Work

The use of LSTM and RF classifiers for forecasting flight-delays at JFK airport with a novel integration of the weather and Holiday data. We recorded an impressive level of 92.2% accuracy, which is relatively higher compared to inspired studies. This study sees improvement by moving from a 4-class classification of delay to a more better 5-class classification for delay.

This research underlines the critical role of weather factors in flight delays, offering valuable insights for the aviation industry to strategic and improve operational efficiency and passenger experience. This study not only contributes to enhancing flight delay prediction but also sets a foundation for future advancements in this area.
Future research should explore advanced feature extraction techniques and diverse LSTM architectures to further refine flight delay classifications. Additionally, integrating real-time data and considering more variables, such as airport operations, could enhance

prediction accuracy. These developments hold promise for significantly improving air traffic management and enhancing the passenger experience.

8 Acknowledgment

It was a great experience working on research based on Flight delay prediction, and I am very grateful for the constant supervision of my mentor Bharat Agrawal who helped me in every step of this project.

References

Addu, S., Ambati, P. R., Kondakalla, S. R., Kunchakuri, H. and Thottempudi, M. (2022). Predicting delay in flights using machine learning, *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, pp. 374–379.

Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., Peterson, E., Sherry, L., Trani, A. and Zou, B. (2010). Total delay impact study, *NEXTOR Research Symposium*, Washington DC.
URL: *http://www.nextor.org*

Borse, Y., Jain, D., Sharma, S. and Vora, A. (2020). Flight delay prediction system, *International Journal of Engineering Research and* **V9**.

Borsky, S. and Unterberger, C. (2019). Bad weather and flight delays: The impact of sudden and slow onset weather events, *Economics of Transportation* **18**: 10–26.
URL: *https://www.sciencedirect.com/science/article/pii/S2212012218300753*

Bureau of Transportation Statistics (2022). Full year 2022 u.s. airline traffic data, `https://www.bts.gov/newsroom/full-year-2022-us-airline-traffic-data`. Accessed: 2023-12-12.

Bureau of Transportation Statistics (2023). Bureau of transportation statistics - on-time departure data, `https://www.transtats.bts.gov/ONTIME/Departures.aspx`. Accessed: December 1, 2023.

Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z. and Zhao, D. (2020). Flight delay prediction based on aviation big data and machine learning, *IEEE Transactions on Vehicular Technology* **69**(1): 140–150.

Jiang, Y., Liu, Y., Liu, D. and Song, H. (2020). Applying machine learning to aviation big data for flight delay prediction, *2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, pp. 665–672.

Kaewunruen, S., Sresakoolchai, J. and Xiang, Y. (2021). Identification of weather influences on flight punctuality using machine learning approach, *Climate* **9**(8).
URL: *https://www.mdpi.com/2225-1154/9/8/127*

Kaggle (2023). Us federal pay and leave holidays, `https://www.kaggle.com/datasets/donnetew/us-holiday-dates-2004-2021`. Kaggle dataset.

Kim, Y. J., Choi, S., Briceno, S. and Mavris, D. (2016). A deep learning approach to flight delay prediction, *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, pp. 1–6.

Lancelotte, A. (2023). Airadvisor study, https://airadvisor.com/en/worst-airports-in-the-united-states-for-delays. Accessed: 2023-11-23.

Li, Q. and Jing, R. (2022). Flight delay prediction from spatial and temporal perspective, *Expert Systems with Applications* **205**: 117662.
URL: *https://www.sciencedirect.com/science/article/pii/S0957417422009642*

Liu, J., Zhao, J., Lu, Y. and Wu, Y. (2022). Flight delay prediction method based on lstm, *2022 International Conference on Automation, Robotics and Computer Engineering (ICARCE)*, pp. 1–4.

Mokhtarimousavi, S. and Mehrabi, A. (2023). Flight delay causality: Machine learning technique in conjunction with random parameter statistical analysis, *International Journal of Transportation Science and Technology* **12**(1): 230–244.
URL: *https://www.sciencedirect.com/science/article/pii/S2046043022000119*

Open-Meteo (2023). Historical weather api documentation, `https://open-meteo.com/en/docs/historical-weather-api`. Accessed: December 1, 2023.

Ye, B., Liu, B., Tian, Y. and Wan, L. (2020). A methodology for predicting aggregate flight departure delays in airports based on supervised learning, *Sustainability* **12**(7).
URL: *https://www.mdpi.com/2071-1050/12/7/2749*

Yiu, C. Y., Ng, K. K. H., Kwok, K. C., Tung Lee, W. and Mo, H. T. (2021). Flight delay predictions and the study of its causal factors using machine learning algorithms, *2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, pp. 179–183.

Yu, B., Guo, Z., Asian, S., Wang, H. and Chen, G. (2019). Flight delay prediction for commercial air transport: A deep learning approach, *Transportation Research Part E: Logistics and Transportation Review* **125**: 203–221.
URL: *https://www.sciencedirect.com/science/article/pii/S1366554518311979*

Zhang, B. and Ma, D. (2020). Flight delay prediciton at an airport using maching learning, *2020 5th International Conference on Electromechanical Control Technology and Transportation (ICECTT)*, pp. 557–560.