Predicting Flight Delays at Dublin Airport: A Machine Learning Approach to Operational and Environmental Impact Analysis

Alexander Bolger

C21449874

Submitted in partial fulfilment of the requirements for the degree of:

BSc in Business Analytics

Technological University Dublin

Supervisor: Ayman Tobail

Submission Date: 01/05/2025

Word Count: 15,420

Declaration

This dissertation is submitted by the undersigned to Technological University Dublin in partial fulfilment for the degree Business Analytics. It is entirely the author's work and has not been submitted previously for an award to this or any other institution. All sources consulted are appropriately referenced as per the TUD School of Business Style Guide.

Signed: Alexander Bolger

Date: 01/05/2025

Acknowledgements

I would firstly like to thank my academic supervisors, Dr. Ayman Tobail and Neil O'Connor, for their continuous help and wealth of knowledge throughout this academic year. To Steven Owen, for sharing his experiences and insights into the aviation industry. To my classmates with whom I have spent the past four years growing together. Finally, to my parents, for their constant support and for all the opportunities they have provided me with.

Abstract

Flight delays represent a significant cost burden for airports, airlines, and stakeholders across the aviation ecosystem. Beyond disrupting operational efficiency and reducing passenger satisfaction, delays also contribute to substantial economic losses and environmental degradation. In 2022 alone, flight disruptions were estimated to cost the aviation sector approximately $67.5 billion USD and resulted in over 9 million tons of CO_2 emissions across the United States, Europe, and Australia alone. This study highlights Dublin Airport, given its status as Ireland's busiest aviation hub and primary international gateway. By leveraging machine learning to forecast delays, the research addresses a critical gap in predictive modelling tailored to airport-specific contexts. The goal is to generate actionable insights that support data-driven decision-making in both operational planning and climate impact mitigation. Ultimately, the framework developed in this study offers a practical tool that can be integrated into real-time airport operations to enhance scheduling, improve passenger experience, and reduce emissions globally.

To address the problem of flight delays, this study develops a machine learning classification framework using historical flight operations and meteorological data from Dublin Airport. The dataset incorporates a range of predictors, including departure time, airline, aircraft type, flight distance and weather conditions such as wind, speed, visibility and humidity. Dublin Airport serves as an exemplar case due to its complex scheduling patterns and exposure to a variety of weather conditions common to many mid-sized international airports. Several models were tested individually, including Random Forest, XGBoost and Multilayer Perceptrons (MLP), with a stacking ensemble model, combining all three, achieving the highest performance. The final model achieved an accuracy of 86% and an AUC score of 0.91. To enhance transparency and trust, SHAP (Shapley Additive exPlanations) values were used to interpret feature importance and enhance trust in the model's predictions.

The model identified key predictors of delay, including temporal factors, flight distance and adverse weather conditions, particularly when such factors interact. Long-haul flights and peak season operations were found to incur the highest economic and environmental penalties. By quantifying excess costs and CO_2 emissions associated with delays, the framework highlights the broader impact of operational inefficiencies. The findings offer a

practical decision-support tool for airport authorities, enabling data-driven improvements in scheduling, resource allocation, and sustainability efforts. Importantly, while developed using data from Dublin Airport, the framework is adaptable to other airports with similar operational and environmental characteristics. Its ability to integrate real-time data and provide interpretable outputs makes it suitable for wider implementation across the aviation sector.

Table of Contents

Contents

List of Figures

List of Tables

Chapter 1 – Introduction

1.1 Introduction

Air transportation plays a crucial role in global infrastructure, enabling the rapid movement of people and goods while allowing for economic growth, international trade, and tourism. As global air traffic continues to grow, so too does the importance of maintaining efficient, reliable, and sustainable airport operations.

For an island nation like Ireland, air travel is a necessity. Dublin Airport is Ireland's primary international gateway and is critical to the country's economy. Handling over 34 million passengers in 2024, DAA (2025) supports trade, tourism, and employment. However, Ireland's coastal location on the edge of the Atlantic Ocean makes it particularly vulnerable to volatile weather conditions. Strong winds, heavy rainfall, fog, and low visibility frequently disrupt operations, leading to flight delays and cancellations.

From an operational perspective, delays reduce schedule reliability and disrupt airport resource planning. Financially, airlines incur additional costs from extended ground handling, increased crew hours, and passenger compensation. Environmentally related to SDG 13 – Climate Action, operational inefficiencies contribute to approximately 6-7% of total aviation fuel burn leading to excess CO2 emissions, ICAO (2022). The global scale of these impacts is estimated at $67.5 billion in delay-related costs and over 9 million tons of CO_2 emissions recorded across the U.S, Europe, and Australia alone in 2022 (AirHelp, 2023). This intersection of operational inefficiency, financial loss, and environmental harm presents a growing challenge for the aviation industry, especially as it faces mounting pressure to follow sustainable initiatives.

While many studies have examined flight delays from an operational perspective, few have comprehensively integrated weather forecasting with predictive modelling of flight delays. Few have quantified the associated financial and environmental impacts, and this represents a key research gap. The lack of predictive frameworks that combine real-time weather data with delay classification, while also estimating the economic and sustainability costs of those delays.

To address this gap, the present study develops a machine learning classification model to predict whether historical flights at Dublin Airport will be delayed, using time-based weather conditions. The model incorporates estimates of both financial loss and carbon emissions resulting from these delays, offering economic and environmental output.

By integrating flight, weather, financial costs and emissions data, the study provides a comprehensive tool for understanding how flight delays affects airport performance. The findings aim to support data-driven decision-making around flight scheduling, risk mitigation, and environmental impact reduction. Ultimately, this research contributes to enhancing the operational improvement and sustainability of Dublin Airport while also providing a scalable approach that could be applied to other airports facing similar challenges.

1.2 Research Question

This research explored the use of machine learning to predict flight delays in the aviation domain. Focusing on Dublin Airport, it aims to assess the broader financial and environmental consequences of those delays. A variety of classification algorithms were applied to historical flight and weather data, emphasising model performance, explainability and real-world applications. Leading to the central question of the study:

"Can machine learning be used to accurately predict flight delays at Dublin Airport, while also providing insights into the financial and environmental impact of disrupted operations?"

1.3 Research Objectives

1. *To investigate the influence of weather-related features on flight delays at Dublin Airport and assess how machine learning models capture and model these effects.*

2. *To evaluate the predictive performance of various machine learning algorithms in forecasting flight delays within a specific airport environment.*

3. *To quantify the contribution of flight delays at Dublin Airport to excess CO2 emissions using aircraft size-related features and operational data.*

4. *To identify and analyse the most influential features contributing to the prediction of flight delays.*

5. *To explore how predictive modelling can inform and support delay mitigation strategies in airport operations.*

1.4 Analysis Framework

This dissertation follows a structured analysis framework, as illustrated in Figure 1. The framework is based on a research, build, and implementation approach, strategically designed to organise each phase of the study. This structure facilitates a seamless transition between stages, ensuring a logical and coherent progression throughout the research process.

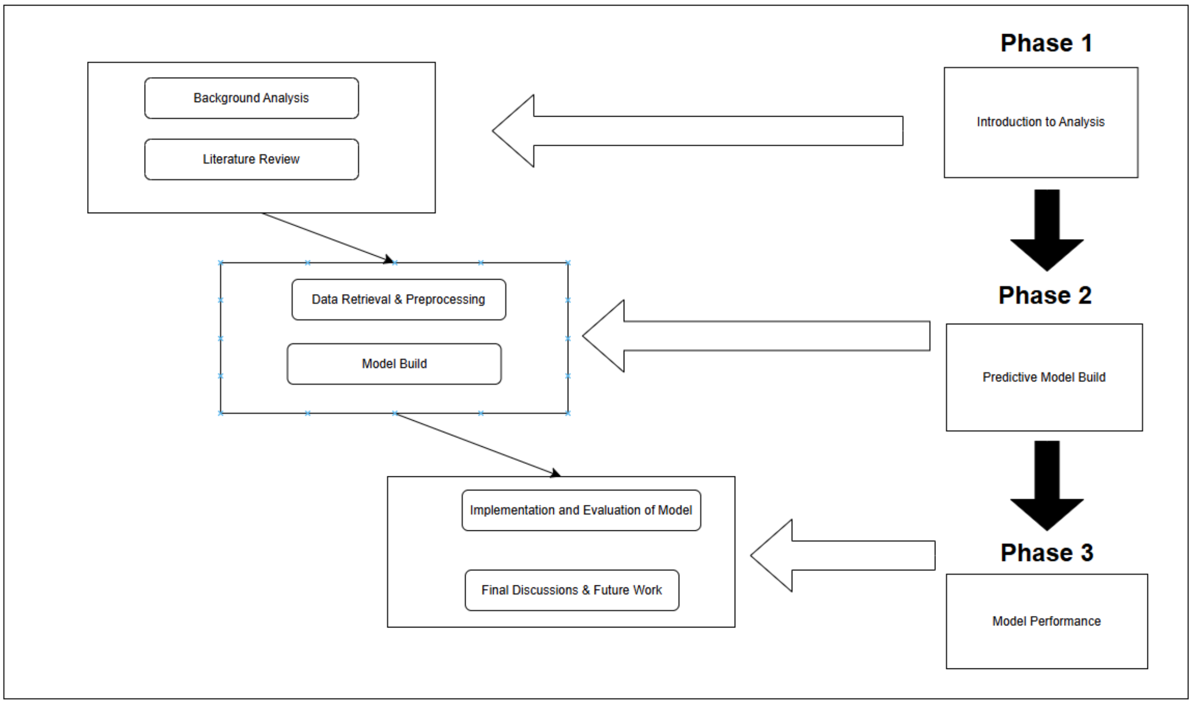


Figure 1 - Analysis Framework

1.5 Structure of Dissertation

This dissertation is organised into seven chapters, each addressing a core component of the research conducted to develop a machine learning framework for predicting flight delays and assessing their operational impact at Dublin Airport.

Chapter 1 introduces the research topic, presenting the motivation, objectives, and central research questions. It also outlines the analysis framework and the overall structure of the dissertation.

Chapter 2 provides a comprehensive literature review, examining previous research on flight delay prediction, the role of weather and operational factors, and the application of machine learning in aviation. It also discusses the financial and environmental impacts of delays and identifies research gaps that this study aims to address.

Chapter 3 outlines the research methodology, including data sources, feature selection, preprocessing steps, and the evaluation metrics used. This chapter also introduces the machine learning models applied, such as Neural Networks, random forest, and gradient boosting methods, along with ensemble techniques like stacking.

Chapter 4 presents the exploratory data analysis (EDA), highlighting key patterns related to temporal factors, weather conditions, operational variables, and their relationship with delay frequency and severity. It also includes an environmental impact analysis based on CO_2 emissions.

Chapter 5 details the model implementation and evaluation process, including hyperparameter tuning, performance comparison across models, and the use of SHAP for model explainability.

Chapter 6 contains the discussion of findings, reviewing the limitations of the study, real-world implications, and the financial and environmental consequences of delays. This chapter also compares outcomes from this study to existing literature. Finally, it revisits the research questions and outlines recommendations for future work.

Chapter 7 concludes the dissertation by summarising the key insights, highlighting the study's contributions to operational decision-making, and emphasising its relevance for improving efficiency and sustainability at Dublin Airport.

Chapter 2 – Literature Review

2.1 Literature Introduction

Weather conditions play a pivotal role in determining the operational efficiency and safety of airport operations. Dublin airport serves as a critical hub, not only connecting Ireland to Europe but also serving as a gateway to the United States. Airports are unique in the way they are extremely susceptible to adverse weather conditions, leading to delays, cancellations and disruptions on the ground and in air traffic management. Research globally has consistently demonstrated that weather is one of the most significant factors affecting airport performance Schultz et al., (2018). The sensitivity of airport operations and strict scheduling to meteorological phenomena makes understanding and mitigating these impacts a priority for enhancing resilience.

For weather analysis, conditions such as low visibility, strong crosswinds, heavy precipitation, and freezing temperatures have been highlighted as key disruptors. Schultz et al. (2018), for example, proposed an algorithmic approach (ATMAP) that uses METAR (Meteorological Aerodrome Reports) data to quantify the operational impact of weather. While this study does not implement the ATMAP framework, it acknowledges its relevance in illustrating how severe weather conditions correlate with increased delays and cancellations. Instead, this study adopts an approach based on raw weather variables to highlight the direct impact of specific meteorological factors.

Several studies have focused on specific weather factors and their influence on the operations. For example, Eurocontrol (2017) highlighted how a reduction in ceiling and visibility, often associated with fog or heavy rainfall can limit the capacity of precision landing systems, causing delays during peak traffic periods. The NASDAC study of NTSB statistics found ceiling and visibility were a contributing factor in 24% of aviation accidents over eight years Kulesa, 2003). Similarly, research has highlighted wind conditions, particularly crosswinds and gusts, disrupting runway operations Kulesa, (2003). This research paper was performed in the US region, looking at the potential disruptions caused by this uncontrollable issue.

2.2 Research Gaps and Opportunities

While significant research has been conducted on the impact of weather conditions at airports, machine learning, financial implications, and environmental consequences of flight delays, notable gaps remain. Much of the existing literature focuses on case studies from large international hubs such as airports in the U.S., Europe, and China (Mtimkulu, (2023) and Hatıpoğlu and Tosun (2024)), with limited research tailored to medium-sized airports like Dublin. Given Dublin Airport's unique operational location, including its reliance on transatlantic routes and a mix of low-cost and full-service carriers, further studies are needed to explore how localised weather patterns and operational constraints influence delay patterns Reitmann, (2019). Additionally, while various machine learning models have been applied to predict delays, there is an opportunity to refine these models by integrating real-time weather forecasting to improve future predictive accuracy. The environmental impact of delays also remains an underexplored area, particularly in quantifying excess fuel consumption and emissions specific to Dublin Airport. Future research should explore the role of optimisation techniques in mitigating these delays while supporting Dublin Airport's carbon neutrality goals. Addressing these gaps will provide more precise, data-driven insights to improve operational efficiency, reduce financial losses, and support sustainability initiatives crucial for sustainable development.

2.2 Machine Learning Methods Used Previously

2.2.1 METAR Data

Aviation Routine Weather Reports (METARS) provide vital surface weather information collected from observation stations located at or near airports. These reports are indispensable for pilots and air traffic management, offering real-time data on key meteorological factors such as wind speed and direction, visibility, cloud cover, temperature, and atmospheric pressure. The decoding and interpretation of this complex data have been extensively studied to support enhanced decision-making in aviation Lui, (2014). METAR data plays a crucial role in ensuring the safety, efficiency, and reliability of airport operations and flight planning Reitmann, (2019). Research on London Gatwick Airport, for example, has demonstrated METAR's strong potential in forecasting disruptions caused by specific weather conditions Reitmann, (2019).

Rather than relying on predefined scoring systems, more recent studies have emphasised data-driven approaches for understanding weather impacts on airport operations Schultz et al. (2019), for example, applied machine learning to link weather conditions directly with airport disruptions, focusing on patterns learnt from real operational data. This approach allows for more flexible and accurate predictions, tailored to the unique conditions of each airport. This approach offers improved adaptability and predictive accuracy, enabling tailored insights for specific airport environments. As this study focuses on optimising operations at Dublin Airport, leveraging raw METAR variables in conjunction with machine learning techniques presents a promising pathway for enhancing delay prediction and improving resource allocation.

2.2.2 Neural Networks for delay and weather classification

Artificial Neural Networks (ANNs) are widely recognised as powerful machine learning techniques, modelled to replicate the functionality of the human brain. They are particularly effective for solving non-linear and stochastic problems due to their advantages such as fault tolerance, handling incomplete knowledge, and parallel processing capabilities (Graupe, (2013); Mijwel, (2018)). ANNs have been shown to outperform other methods in capturing complex, non-linear relationships in data, making them an ideal choice for delay and weather classification (Venkatesh, (2017); Thiagarajan, (2017)). These attributes provide a comparative advantage when integrating diverse datasets, such as weather, operational, financial, and environmental data, offering a holistic view for more accurate predictions.

Studies have demonstrated that ANNs often achieve higher classification accuracy compared to other machine learning techniques. For instance, Purushothaman (2024) found that ANNs outperformed ARIMA models, achieving an accuracy of 86.90% compared to 85.94%. This highlights their capability to handle complex classification tasks effectively.

However, the performance of ANNs heavily depends on the optimisation of hyperparameters, including the number of layers, learning rate, dropout rate, activation functions, and training duration. Arya (2017) emphasised the volatility in results when varying the number of layers and neurons, which led to a range of 22% in performance outcomes. This underscores the importance of careful tuning and testing of hyperparameters to achieve optimal results.

In this research, a specific form of ANN, the Multi-Layer Perceptron (MLP), was used for flight delay classification at Dublin Airport, given its strong ability to model complex non-linear interactions across operational and weather-related features.

2.2.3 Ensemble techniques used for classification

Ensemble techniques are very common throughout research surrounding classification flight delay models in airports. A stacking approach combines multiple models to improve predictive performance. Mtimkulu (2023) highlighted the superior performance of an ensemble approach, achieving an impressive accuracy rate of 92.4%. This suggests that stacking's ability to combine diverse base models effectively captures the complex patterns present in flight data. The Random Forest algorithm also showed strong predictive capabilities, with an accuracy of 91.2%.

In comparison, an interesting discovery was a very low performance in AdaBoostClassifier only achieving an accuracy of 51.6%. These studies are crucial to highlight key features influencing flight delays. In the same domain, a study done by Wang (2022), in Beijing airport displayed strong results for stacking algorithms outperforming SVR, LGBM, RF, AdaBoost, KNN and Logistic Regression.

An advantage of an ensemble approach can be seen in a paper by Schwarz (2022) as he demonstrated that a stacking ensemble of multiple base learners could provide more accurate predictions of commercial flight delays compared to individual models. This paper highlights the use of LASSO, random forests and neural networks as base learners in the stacking ensemble.

When comparing results, a study by Wang et al. (2022) used gradient boosting ensemble models to build machine learning flight delay prediction model, this study captured promising results displaying the various factors influencing delays. On the flip side, a comparative study by Tang (2021) evaluated seven different machine learning algorithms for flight delay prediction, including ensemble methods like random forest and gradient boosted trees. His results showed how tree-based ensemble classifiers generally outperformed other base classifiers. These two comparison studies display the need for trial and error amongst base classifiers to achieve the best outcome possible.

2.2.4 Applicability to Dublin Airport

Dublin Airport, as a major hub for transatlantic flights and a mix of low-cost and full-service carriers, faces operational challenges influenced by weather variability and scheduling complexities. Studies such as Reitmann (2019) at London Gatwick have demonstrated the effectiveness of using METAR data to forecast weather-driven disruptions. Ensemble methods, particularly stacking, offer a promising solution for delay prediction. Research by Mtimkulu (2023) and Wang (2022) has shown that stacking outperforms individual models by combining diverse base learners to capture complex flight delay patterns. Given Dublin Airport's unique characteristics, its role as a transatlantic hub, diverse weather conditions, dual usage from low cost and full-service airlines leading to operational complexity, stacking could provide more accurate and robust delay forecasts. By integrating METAR data with ensemble learning, Dublin Airport could enhance operational efficiency, minimise disruptions, and improve passenger experience.

Table 3 Summary of prior prediction of flight delay

	Sources	Method	Datasets	Data period	Delay time (min)	Results
Machine learning	Khaksar and Sheikholeslami [9]	Bayesian modeling, decision tree, cluster classification, random forest, and hybrid method	US and Iranian airline	US: 6 months, Iran: 16 months	0–15, 15–60, 60+	Accuracy more than 70%
	Al-Tabbakh et al. [11]	Decision tree, random forest, and REPTree	Egypt airline	Jan 2018 (1 month)	–	Accuracy around 80.3%
	Ye et al. [12]	Multiple linear regression, support vector machine, extremely randomized trees, and LightGBM	Nanjing Lukou airline	Mar 1st 2017 to Feb 28th 2018	15+	Accuracy of 86.53%
	Atlioglu et al. [13]	11 machine learning models. CART, KNN, GBM, XGB, and LGBM	Dammam King Fahd International Airport	Jan 1st 2017 to Dec 9th 2019	15+	Accuracy around 82%
Neural network	Kim et al. [8]	LSTM, RNN	ATL, LAX, ORD, DFW, DEN, JFK, SFO, CLT, LAS, PHX	Jan 2010–Aug 2015	15+, 30+	Accuracy of 90.95%
	Qu et al. [10]	CBAM-CondenseNet and SimAM-CNN-MLSTM	The Civil Aviation Administration of the China East China Regional Administration (ECRA)	Mar 2018–May 2019	15–60, 60–120, 120–240, 240+	Accuracy of 89.8%, 91.36%
	Yazdi et al. [14]	Stack denoising autoencoder- levenberg marquart model, SAE-LM, SDA	The Bureau of Transportation Statistics of United State Department of Transportation	For 5 years	15+	Accuracy of 96%, 86%, 89%

Figure 2 - Previous ML Literature Results Kim, (2024)

2.3 Impact Around the World

Many studies have been conducted globally, highlighting the financial effects of flight delays. There are several expenses associated with flight delays, some of which are direct, such as increased fuel consumption and crew costs, and some of which are indirect, including reduced customer satisfaction and lost productivity Anupkumar, (2023). These costs are experienced not only by airlines but also by passengers, businesses, and broader economies. According to a study by the Federal Aviation Administration (2020), the US economy lost up to $32.9 billion due to flight delays in 2019. These costs include both direct financial costs and indirect economic impacts, such as missed business opportunities and decreased productivity. The study emphasises that delays not only disrupt air traffic but also have widespread consequences on the economy.

A similar trend has been observed in Europe, where delays have led to significant financial losses. A survey conducted by the Global Travel & Tourism Council (2018) estimated that Europe suffered a €9.6 billion loss due to flight cancellations and delays (World Travel and Tourism Council, 2019). These losses were attributed to decreased economic output, lost business opportunities, and reduced visitor spending. Efthymiou et al. (2019) found that delays at major European hub airports, such as London Heathrow, lead to substantial disruptions in airline schedules, negatively impacting customer satisfaction and airline profitability. Their study highlighted that passengers increasingly expect delays, which, while fostering resilience, also reduces confidence in airline services.

Similarly, the economic impact of flight delays has been studied in other regions with significant reliance on tourism. For instance, a study conducted in New Zealand, a country with an aviation sector like Ireland's, identified how flight delays influence local businesses and tourism. Brickell and Alberts (2016) found that flight delays led to a decline in customer satisfaction and an estimated NZD 56 million loss in tourism revenue. The study underscores the importance of reliable airline operations for small, tourism-dependent economies, where delays can disrupt travel plans and deter repeat visitors.

2.3.2 Financial Loss for Airlines and Passengers

Airlines are responsible for compensating passengers for delays, rebooking flights, and covering accommodation expenses when necessary. However, there are also hidden costs that may not be immediately apparent to passengers. These include increased fuel consumption, higher airport fees, additional crew expenses, and maintenance costs associated with extended turnaround times Anupkumar, (2023). According to Johnson (2019), on average, each delayed flight in the US costs airlines approximately $8,000, with longer and international flights incurring even greater expenses.

The financial burden of delays extends beyond airlines to airports. Increased operating costs arise from the need for additional staff, maintenance services, and heightened security measures to accommodate stranded passengers Brown., (2020). Airports also suffer revenue losses when passengers, frustrated by repeated delays, opt to fly from alternative airports in future travel plans. According to Efthymiou et al. (2019), congestion at hub airports like Heathrow exacerbates financial losses, as cascading delays disrupt operations across entire airline networks.

Passengers too, experience financial losses due to flight delays. Beyond the inconvenience, delays lead to additional out-of-pocket expenses such as hotel stays, meals, and alternative transportation arrangements. According to a report by the US Department of Transportation (2020), flight delays cost passengers an estimated $2.4 billion in lost time and out-of-pocket expenses in 2019. Furthermore, frequent delays contribute to passenger dissatisfaction, which can ultimately impact airline brand loyalty and future revenue streams Baranishyn et al., (2010).

2.3.3 Customer Satisfaction and Management of Flight Delays

Customer satisfaction and loyalty are critical for airlines, as these factors contribute to long-term profitability. Extensive research has explored the impact of flight delays on passenger satisfaction, emphasising the role of delay management in shaping consumer perceptions of service quality. One key strategy for mitigating passenger dissatisfaction during delays is improving communication regarding delay duration and expected resolution times Baranishyn et al., (2010). Their study found that timely and accurate information significantly reduces passenger anxiety and fosters a more positive attitude toward the airline.

Tarmac delays pose unique operational challenges for airlines and airports. Passengers often feel trapped during extended tarmac delays, which amplifies frustration and dissatisfaction. Baranishyn et al. (2010) analysed passenger experiences during tarmac delays and found that both the timing and accuracy of delay information influence consumer perceptions. Their findings indicate that airlines that provide timely updates and realistic expectations regarding the delayed duration are more likely to retain customer loyalty. In contrast, inaccurate or delayed communication can exacerbate frustration, leading to long-term reputational damage for the airline.

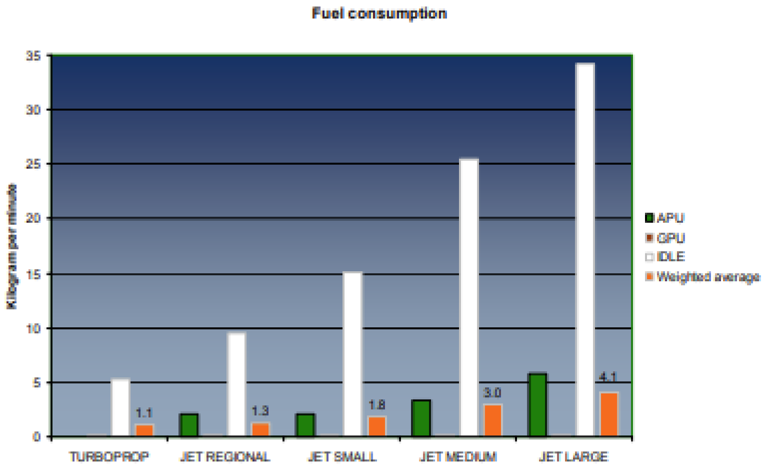
Research has also highlighted the financial consequences of poor delay management. A study by Efthymiou et al. (2019) found that flight delays and cancellations remain the leading cause of passenger complaints, contributing to a 9.5% decline in customer satisfaction in the airline industry over the past decade. Moreover, their findings suggest that 72% of passengers notice a delay within the first 10 minutes of stopping on the tarmac, underscoring the importance of rapid communication. Ultimately, addressing the financial and customer satisfaction impacts of flight delays requires a combination of improved operational efficiency, enhanced communication strategies, and policy interventions. As airlines and airports continue to navigate these challenges, further research is needed to develop effective mitigation strategies that balance cost reduction with passenger experience improvements.

2.4 Environmental Impacts of Flight Delays

Flight delays have been extensively studied in aviation research due to their economic, social, and environmental implications. While much attention has been given to passenger inconvenience and operational costs, less emphasis has been placed on their environmental consequences. EuroControl, (2008) highlights the environmental impact of delays, attributing excess emissions to fuel consumption during idle times, rerouted flights, and extended usage of auxiliary power units (APUs) seen in FIG 3. Their study estimates that during delays, APUs contribute approximately 9% of fuel consumption, ground power units (GPUs) 81%, and idle engine operations 10%. These figures are calculated as weighted averages across diverse aircraft sizes, underscoring the variability of environmental impacts based on fleet composition.

Research on mitigating these effects often focuses on operational efficiency. Dublin Airport's sustainability initiatives, for example, align with international climate objectives, aiming for net zero carbon emissions by 2050 (Dublin Airport Carbon Reduction Strategy, 2021). Specific measures include a 50% reduction in the carbon footprint by 2030 and achieving ACI Level 4+ accreditation for carbon neutrality (Environmental Sustainability Policy, 2019).

However, while these goals are ambitious, there is limited research evaluating their feasibility or real-world effectiveness, particularly in managing emissions during delays.



Several studies have identified gaps in understanding and measuring the environmental impact of flight delays at airports with varying infrastructure capabilities and operational characteristics. Existing research often focuses on individual airport case studies or generalised industry trends, leaving a lack of comprehensive, comparative analyses across diverse contexts (EuroControl, 2008; Dublin Airport Carbon Reduction Strategy, 2021). This gap in the literature highlights the importance of further research into scalable, data-driven solutions that can optimise ground operations, improve fuel efficiency, and mitigate the environmental footprint associated with delays. Addressing these challenges is particularly crucial as airports strive to align with global sustainability goals, such as achieving carbon neutrality and reducing greenhouse gas emissions.

2.5 Resilience Strategies for Flight Delays

2.5.1 Operational and Technological Solutions

Many studies have examined methods to improve efficiency and reduce flight delays in aviation. One widely researched approach is the implementation of collaborative decision-making (CDM) systems. These systems enable real-time data sharing among airlines, airports, and air traffic control authorities, enhancing operational coordination and efficiency. Odoni and Belobaba, (2017) highlight that CDM systems can significantly reduce airport delays and improve overall performance. Another key strategy involves upgrading air traffic control (ATC) systems to reduce congestion and enhance airport operations. Zhang and Zhao, (2021) demonstrated that integrating advanced technologies, such as Automatic

Dependent Surveillance-Broadcast (ADS-B) and CDM, improves air traffic management (ATM) system performance, leading to a decrease in passenger flight delays.

An emerging strategy in airport operations is the introduction of terminal competition, aimed at encouraging efficiency and innovation through direct rivalry between terminals. This approach has been driven by government policy initiatives as a means of addressing capacity constraints and reducing delays. Mclay and Reynolds-Feighan, (2005) explored this concept in Dublin Airport, by fostering competition, there was an improvement in resource allocation, optimisation of gate usage and enhanced passenger flow. These measures contribute to the overall resilience of airport operations by reducing bottlenecks and improving adaptability during disruptions and flight delays.

2.5.2 Infrastructure Developments

Developing airport infrastructure is a widely recognised strategy for reducing delays by increasing capacity and decreasing waiting times. This often involves constructing new terminals, improving gate distribution, and enhancing ground-handling operations Adigun & Adebiyi, (2021). Such initiatives are critical for addressing growing passenger demand and mitigating congestion.

Recent examples of infrastructure improvements underscore their potential impact. Melbourne Airport, for instance, has commenced a $3 billion (AUD) project to construct a third runway, which is projected to contribute an additional $6 billion (AUD) annually to the state economy and create 51,000 jobs. This development has garnered support from aviation experts, such as RMIT University's Chrystal Zhang, who emphasise its potential to significantly reduce delays (Melbourne Airport Runway Project, 2023). The success of such large-scale infrastructure investments highlights their role in improving operational efficiency, but further research is needed to evaluate their scalability and effectiveness across airports of varying sizes and capacities.

2.5.3 Environmental and Collaborative Approach

There are many environmental and collaborative approaches that come hand in hand with innovative solutions, these innovations would aid in the mitigation of environmental loss due to flight delays.

Sustainable Aviation Fuels (SAF) can be implemented nationwide as production increases and adoption through partnerships between airlines, fuel producers and airports. A study is currently underway, conducted by UL (University of Limerick) and TCD (Trinity College Dublin) (https://www.ul.ie/news/ul-and-tcd-to-collaborate-on-sustainable-aviation-fuel-research-study) as they focus on developing SAF production in Ireland, to meet EU mandates for SAF usage by 2050. A study conducted by Shehab (2023) highlighted the potential of the EU to meet the SAF targets in 2030 and 2050. The EU can meet its SAF mandates soon, but long-term success depends on overcoming certain limitations and implementing new technologies.

In addition to SAF initiatives, airports are adopting green airport concepts, which aim to focus on resource conservation, environmental friendliness and efficient operations to minimise environmental impacts. Li et al (2022) explored the concept of green airports, which integrates sustainability with airport operations by promoting energy conservation, emission reduction and the use of renewable resources. Li et al (2022) showed how airports such a Beijing Daxing International Airport and Kunming Changshui International Airport have successfully implemented sustainable strategies incorporating clean energy technologies, waste recycling programs and efficient logistics to reduce flight delays and operational inefficiencies.

Dimitiou and Karagkouni (2022) introduced an evaluation framework for airport sustainability strategies, emphasising the importance of environmental awareness in managing large transport hubs. They highlighted how emissions, noise and energy management and waste reduction initiatives contribute to a more resilient and sustainable aviation sector. Dimitriou and Karagkouni (2022) is crucial for Dublin Airport's development as it displays real life examples in Copenhagen and London Stansted of how operational efficiency has been maximised through collaborative initiatives with local governments and industry stakeholders

2.7 Summary of Literature Review

Various methods and approaches have been explored to analyse flight delays, each providing different levels of accuracy and effectiveness. The literature highlights key factors influencing delays, such as adverse weather conditions, financial implications, and environmental impacts. Studies have introduced advanced machine learning techniques, including neural networks and ensemble methods, to predict delays more accurately using data like METAR reports. Research also emphasises the substantial economic losses caused by flight delays, affecting airlines, passengers, and overall airport operations. Additionally, the environmental impact of delays, including increased emissions and energy consumption, has been examined, highlighting the importance of sustainable initiatives such as green airport concepts and Sustainable Aviation Fuels (SAF). Collaborative approaches between airlines, airports, and policymakers have been proposed to enhance operational efficiency and resilience. The reviewed studies provide a comprehensive understanding of flight delays and serve as a foundation for further research into optimising airport performance and mitigating disruptions.

Chapter 3 – Research Methodologies

3.1 Data Collection

Data plays a central role in airport operations, arriving in various forms that offer diverse analytical opportunities. In this project, two primary sources of data were utilised: flight operations data and meteorological data, each collected via different programming methods.

FLIGHT DATA via REST API

The flight operations data was sourced from Aviationstack via its subscription-based REST API. The API allowed flexible querying through specific parameters to retrieve only necessary information. As this project is focused on Dublin Airport, the ICAO airport tag 'EIDW' was used as a filter by airport. The data collection spanned one year, starting from 01/03/2024 and ending on 28/02/2025.

HTTP requests were used to enable the seamless integration, and the responses were returned in JSON format. This allowed Python to retrieve the data using the REQUESTS library and parsed using JSON. This was then transformed into a structured tabular form using PANDAS before exporting it to a CSV. The CSV format allowed for seamless integration with external weather data.

Weather Data via Web Scraping

Weather data was collected from ogimet using web scraping techniques, with observations recorded at 30-minute intervals for the entire year to match the granularity of flight events.

The process involved sending an HTTP GET request to retrieve METAR reports for a specific month, followed by parsing the HTML content using BEAUTIFULSOUP to extract the raw METAR strings. These METAR reports were then processed using the PYTHON Metar package, which systematically decomposes each string into its meteorological components, such as station ID, observation time, wind speed, visibility, temperature, dew point, and atmospheric pressure, humidity. This process was repeated 12 times, once each month, and the resulting datasets were combined to create a comprehensive year-long weather record for further analysis.

For instance, the following METAR string:

METAR EIDW 011200Z 25006KT 190V290 9999 OVC016 12/09 Q1026

was parsed and transformed into a structured row with each meteorological component extracted and labelled. The results of each parse were stored in individual rows to be used later during preprocessing. A flowchart of the process can be seen in FIG 4 of the process.

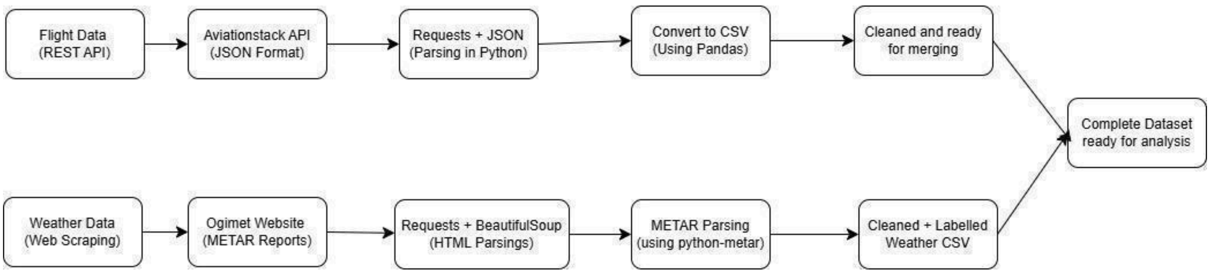


Figure 4 - Dataset Formation Process

3.2 Dataset and Features

The dataset spans a full calendar year, beginning in March 2024 and concluding in February 2025, comprising exactly 365 days of flight operations. Each departing flight from Dublin Airport within this period was matched with the corresponding weather conditions recorded at 30-minute intervals, providing an aligned dataset for analysis. To maintain a manageable dataset size, a random sample of 100 departing flights was selected from each day. This approach, shaped partly by rate limits on the Aviationstack API, provided sufficient variation across seasons, airlines, and aircraft types to support the research objectives.

This time frame coincides with increased public and media focus on Dublin Airport's sustainability and operational future, adding broader relevance to the study. The primary objective of this project is to develop a classification model where the target variable is a binary outcome indicating whether a flight was delayed by more than 15 minutes beyond its scheduled departure time. According to the Federal Aviation Administration (FAA), a delay of 15 minutes or more is typically used as the standard threshold to classify a flight as delayed (FAA, 2024). This benchmark was adopted for this analysis.

The final dataset used for machine learning consists of 29,887 individual flights (rows) and 21 variables, encompassing both flight-specific and weather-related features. Preliminary feature engineering was applied to derive new variables aimed at improving model performance, as seen in the next section. The full modelling dataset is visible in Appendix A.

3.3 Data Preprocessing

To ensure the quality, reliability, and suitability of the dataset for predictive modelling, a structured series of preprocessing steps was implemented. When collecting flight information via the REST API, several columns were found to be irrelevant to the research objectives, such as arrival details, aircraft identifiers, and flight status. These and other related fields were removed to reduce dimensionality, simplify the dataset, and eliminate noise that could negatively affect model performance or the targeted outcome.

The dataset contained a substantial number of missing values, which were addressed through a combination of imputation and row elimination. Categorical fields such as Departure Gate and Departure Terminal were imputed using the mode, assuming consistency in airport operations. Rows missing critical information, such as the Actual Departure Time, were excluded to preserve data integrity and prevent further errors.

To ensure consistency in aircraft data, missing values in the Aircraft Model column were filled using a manually curated airline-to-aircraft mapping, based on known fleet configurations. This enabled the creation of a derived numerical feature, Aircraft Size, which categorised aircraft into size classes (small, medium, large, extra-large). This feature facilitated analysis of the potential influence of aircraft size on delays and was also used in the sustainability and economic focused exploratory analysis examining CO_2 emissions and excess costs per minute of delay, as discussed in Section [4.5].

To prepare the full dataset for machine learning, label encoding was applied specifically to the ‘Airline’ feature. Airline names were converted into numeric codes to ensure compatibility with algorithms requiring numerical input.

Further preprocessing involved the construction of a binary target variable, Delayed, where flights with a departure delay greater than 15 minutes were assigned a value of 1 (delayed), and others a value of 0 (on time). In addition, several temporal binary features were engineered to capture operational stressors and patterns, including peak hour, busiest day, and peak season, each reflecting time-based trends known to influence delays, discussed further in Section [3.4].

To ensure consistent feature scaling and enhance the performance of distance-based and gradient-based models, all continuous numerical variables were standardised using z-score normalisation via StandardScaler from the scikit-learn library, transforming them to have a mean of zero and a standard deviation of one.

Finally, the preprocessed dataset was partitioned into a training set (80%) and a test set (20%). The training set was used to fit the predictive models, while the test set was held out for evaluating model performance on unseen data and ensuring generalisability.

3.4 Feature Engineering

Feature engineering is a crucial step in preparing the dataset for predictive modelling. The aim was to transform raw weather and flight-related data into meaningful features that could better capture patterns behind flight delays at Dublin Airport. This process helps improve model accuracy by reducing noise and emphasising relevant operational conditions which are most likely to influence delays.

Some of the key engineered features included:

- **Day of the Week**: According to EUROCONTROL (2024), the busiest days at European airports are Friday, Sunday, and Monday. To incorporate this, a binary variable was created to indicate whether a flight occurred on one of these high-traffic days (1) or not (0).

- **Busy Seasons**: EUROCONTROL (2024) also identified June through October as the busiest period for airport operations. A seasonal binary variable was added to reflect whether a flight took place during this peak window.

- **Rush Hour at Airport**: Keenon (2013) highlighted that the busiest times at European airports are typically from 6 a.m. to 9 a.m. and again from 4 p.m. to 7 p.m. A binary variable was added to identify whether a flight occurred during these rush hours (1) or outside of them (0).
- **Delay Label**: A binary target variable was created, where a flight was labelled as 'delayed' if the departure delay exceeded 15 minutes. This was created to be the target variable of the analysis. This helped simplify the prediction problem and align it with industry standards.

3.5 Evaluation Metrics

Evaluation metrics are critical for assessing the performance of the various predictive models. As defined by Brownlee (2021), an evaluation metric "quantifies the performance of the predictive model". There are specific classification based metrics used in this analysis:

- Accuracy
- Precision
- Recall / Sensitivity
- F1 Score
- AUC / ROC

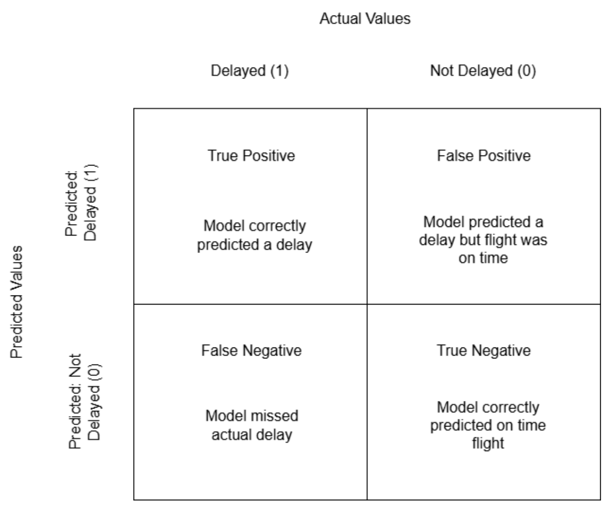


Figure 5 - Classification Confusion Matrix

Accuracy

Accuracy for this study looks at the proportion of flights that are correctly classified (both delayed and not delayed) among all of the flights in the dataset. This provides an overall view of the model's performance but caution must be taken when classifying imbalanced datasets, due to poor performance on minority classes.

$$\text{Accuracy (acc)} \qquad \frac{tp + tn}{tp + fp + tn + fn}$$

Figure 6 - Accuracy Formula Hossin (2015)

Precision

Precision in this analysis quantifies the proportion of true positives among all flights the model predicted as delayed. It focuses on how accurate the model is when it predicts a delay.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

Recall

Recall in this study measures how many of the actual delayed flights the model correctly predicted. It shows how good the model is at finding delayed flights, crucial as missing a delay is costly.

$$Recall\ (Sensitivity) = \frac{TruePositive}{TruePositive + FalseNegative}$$

Figure 8 - Recall Formula Muraina (2023)

F1 Score

The F1 score is a combination of precision and recall into one figure. It helps show how well the model is performing overall, especially when the dataset is imbalanced. The higher the F1 score the better the performance.

$$F1 = 2 * \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}}$$

Figure 9 - F1 Score Formula Muraina (2023)

AUC / ROC

The ROC curve shows how well the model is performing when asked to separate the delayed and non-delayed flights by comparing the true positive rate to the false positive rate at different thresholds. The AUC (Area Under the Curve) gives one value to sum up the performance. The closer to 1, the better the model. It is a crucial tool for evaluating model performance.

True Positive (TP)	False Positive (FP)
False Negative (FN)	True Negative (TN)

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

$$FPR = \frac{FP}{FP + TN}$$

Figure 10 - Confusion Matrix - AUC calculation

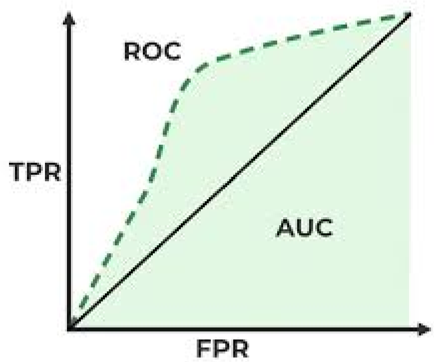


Figure 11 - ROC curve displaying model performance with AUC

3.6 Machine Learning Models

Predictive modelling leverages historical and current data to generate insights and forecast future outcomes by identifying and extrapolating patterns. Through machine learning, algorithms are trained on experience data, enabling the resultant model to make informed predictions on new, unseen instances. For example, by learning from data on previously observed watermelon ripeness indicators, a model can accurately predict the ripeness of an uncut watermelon. According to Zhou (2021), if we consider computer science as a discipline focused on algorithms, machine learning specifically concentrates on algorithms that learn from data.

In the context of this research, machine learning techniques analyse historical flight and weather data combinations to detect patterns that can predict future flight delays. The primary objective was to develop a predictive model capable of identifying the impact of weather conditions on potential flight delays and assessing their environmental impacts. This insight allows the Dublin Airport Authority (DAA) to proactively manage and mitigate these disruptions. The target variable utilised, labelled 'Delayed', is discrete-binary, meaning it possesses only two possible outcomes: delayed (1) or on-time (0). Given this binary nature, classification algorithms are most suitable for this research, as they are specifically designed to predict categorical outcomes. Classification is recognised as a fundamental approach within supervised learning, offering efficient methods to handle categorical data and effectively predict group membership Soofi & Awan, (2017). In this study, the categorical outcomes are explicitly labelled as ‘Delayed’ (1) and ‘On-time’ (0). The accuracy of the model in predicting a flight's departure status directly correlates with the overall performance of the model.

This chapter explores various machine learning algorithms evaluated before constructing the predictive model. Each algorithm underwent assessment based on specific evaluation metrics designed to measure its predictive performance, explaining why certain algorithms demonstrated superior accuracy and effectiveness over others.

3.6.1 Logistic Regression

Logistic regression is a powerful statistical method often used to model the relationship between one or more variables and a binary outcome, such as whether a flight is delayed or on time. It is commonly used as a baseline model for comparison with more complex approaches. As LaValley (2008) explains, the coefficients in a logistic model can be exponentiated to produce odds ratios, which show how the odds of the outcome change with a one-unit increase in a predictor variable. Sperandei (2014) further emphasises the method’s ability to incorporate both continuous and categorical variables, making it particularly useful for research like this, where multiple factors are at play. One of the main advantages of logistic regression is that it provides clear, interpretable results. However, it's important to note that even a large odds ratio doesn’t always imply a high probability of the event occurring.

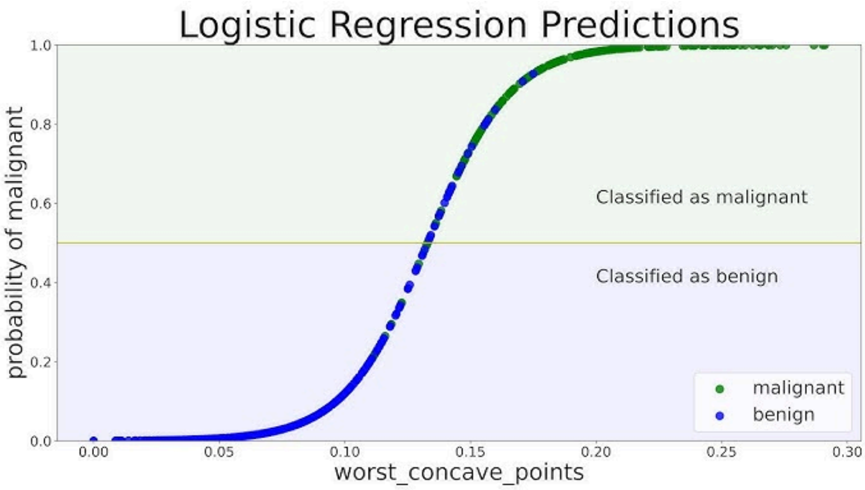


Figure 12 - Logistic Regression Visualisation

3.6.2 Random Forest

The Random Forest algorithm, introduced by Breiman (2001), is a highly popular supervised learning technique based on ensemble learning methods. It is an adaptable model that can be effectively utilised for both classification and regression tasks within machine learning. Random Forests are composed of multiple decision trees, with each tree being constructed independently using random subsets of the available data and selected features. This random selection introduces significant variability among the trees, reducing the correlation between them and enhancing the robustness of the model. During the tree-building process, each node is split based on criteria such as Gini impurity or entropy, both of which measure the purity of a node. The specific splitting criterion is typically determined during the hyperparameter tuning phase to optimise model performance. Other critical hyperparameters include the number of trees, maximum tree depth, minimum samples required per leaf, and the number of features considered at each split. A notable strength of the Random Forest algorithm lies in aggregating predictions from this ensemble of diverse trees, commonly achieved through majority voting for classification problems Breiman (2001). This aggregation effectively reduces overfitting and enhances generalisation, making Random Forests very applicable to a study in this domain.

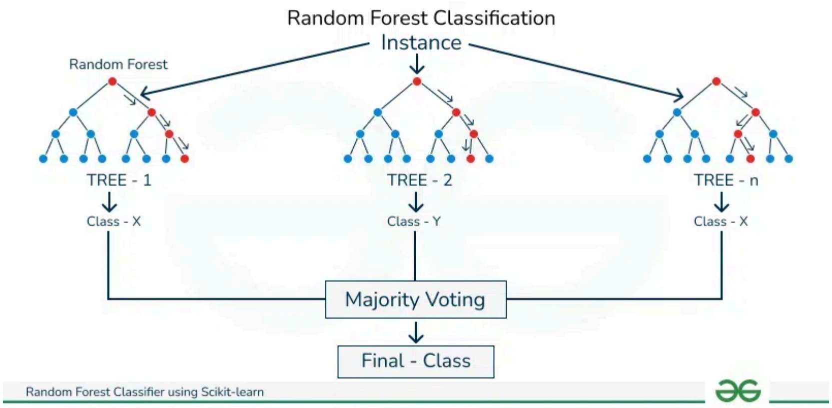


Figure 13 - Random Forest Visualisation

3.6.4 XGB (Extreme Gradient Boosting)

XGBoost (Extreme Gradient Boosting) is a powerful ensemble algorithm built on the principle of boosting, widely used for both classification and regression tasks due to its high predictive accuracy and computational efficiency. Developed by Tianqi Chen, it significantly enhances traditional gradient boosting by optimising for speed, often delivering performance up to ten times faster than standard implementations Chen, (2016). Its ability to handle large and complex structured datasets, combined with support for Python, R, and other languages, has made it a go-to algorithm in real-world applications.

XGBoost builds trees sequentially, with each tree attempting to correct the prediction errors of the previous one. This iterative learning approach, along with hyperparameter tuning (e.g., learning rate, tree depth, and regularisation), allows it to balance bias and variance effectively and avoid overfitting. In this project, these qualities are especially valuable given the mix of weather and flight-related features with non-linear interactions.

A major advantage of XGBoost is its built-in support for feature importance and SHAP values, which provide transparency by explaining individual predictions. This interpretability is crucial in aviation, where understanding the factors driving delays is as important as accurately predicting them. Studies such as Huber et al. (2022) have demonstrated that XGBoost can match or even outperform deep learning models on structured data tasks while using less computational resources and offering greater transparency, making it particularly suitable for this flight delay classification problem.

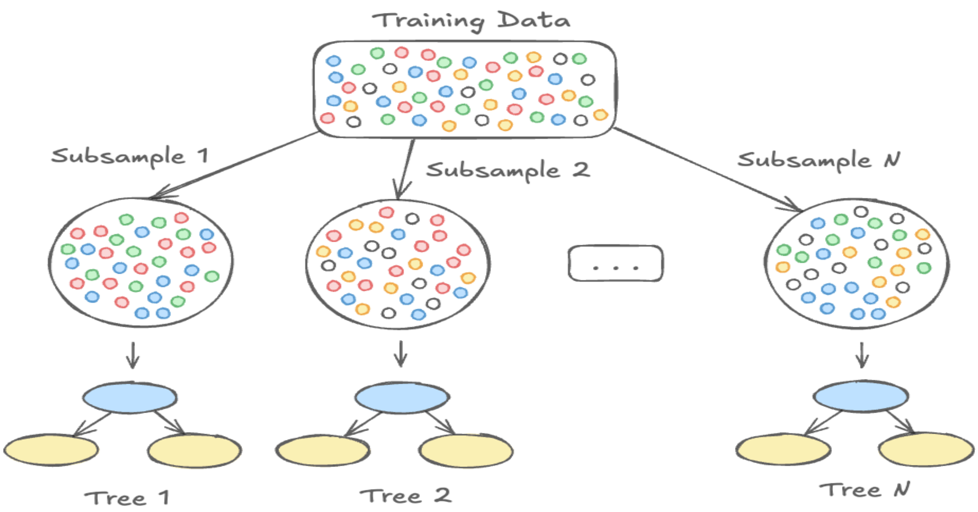


Figure 14 - XGBoost Visualisation

3.6.5 SVM

Support Vector Machines (SVMs) are powerful and reliable machine learning algorithms, especially effective for classification tasks. They work by identifying the optimal boundary that separates different classes, in this case, delayed vs. on-time flights, by maximising the margin between them. This improves classification accuracy and reduces overfitting, even when dealing with high-dimensional or complex datasets.

A key advantage of SVMs is their use of kernel functions, such as the Radial Basis Function (RBF), which enables them to handle non-linear data by projecting it into a higher-dimensional space. This flexibility makes them well suited for real-world datasets with a mix of numerical and categorical variables. According to Khan et al. (2024), among six kernel types tested, the RBF kernel consistently outperformed others in classification accuracy, demonstrating its strength in managing non-linear relationships.

In the context of this project, where flight delays are influenced by numerous interdependent factors like weather conditions, flight schedules, and operational variables, SVMs are particularly valuable. These features often interact in non-linear and unpredictable ways, which SVMs are designed to model effectively. The ability of SVMs to generalise well and focus only on the most critical data points (support vectors) also contributes to efficient training and greater interpretability, important factors in aviation contexts where decision-making needs to be both data-driven and explainable.

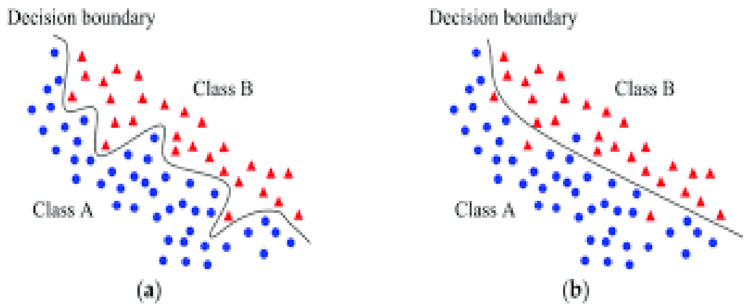


Figure 15 - SVM visualisation

3.6.6 Neural Networks (Deep Learning)

Neural networks, especially Multi-Layer Perceptrons (MLPs), are powerful supervised learning models designed to uncover complex patterns and relationships within data, making them particularly effective in the aviation domain. MLPs consist of multiple layers of neurons and non-linear activation functions, enabling them to capture intricate dependencies between features. These networks follow a feed-forward architecture, where data flows in one direction from input to output without any cycles. Studies by Popescu et al. (2009) highlight how MLPs leverage non-linear activation functions such as sigmoid and tanh to model highly complex relationships. This is further supported by Naskath et al. (2023), who demonstrated the effectiveness of MLPs in handling high-dimensional and uncertain datasets, similar in structure to the one used in this study. These characteristics make MLPs a powerful tool for capturing non-linear feature interactions and generating reliable predictions in challenging environments like flight delay forecasting.

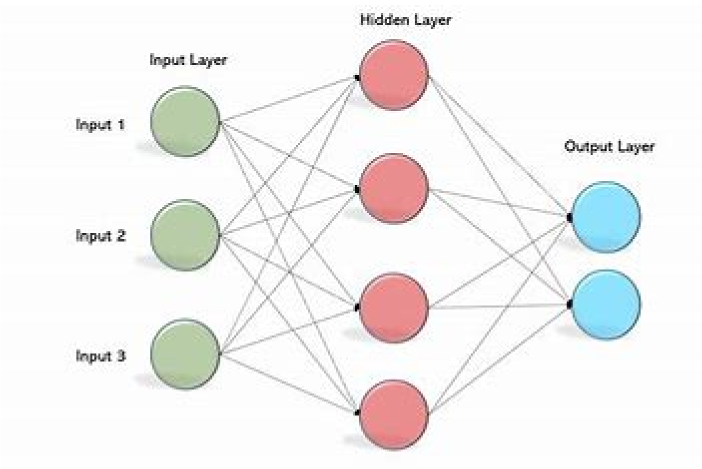


Figure 16 - Neural Network Visualisation

3.6.7 Ensemble Methods - Stacking

Ensemble and stacking methods are advanced machine learning techniques that combine the strengths of multiple individual models to improve the overall prediction accuracy and robustness of research. As we talked about individual models before this, ensemble methods combine these models by aggregating predictions from several models to reduce variance, bias and the risk of overfitting. Stacking, located in the ensemble family, is a more sophisticated strategy that involves training a meta-model to learn how best to combine the outputs of diverse base models (SVM, Random Forest, Logistic Regression, etc.) into a final optimised prediction.

One key advantage to this method is its ability to leverage strengths from models while mitigating individual weaknesses. Research done by Yi, (2021) applied a stacking ensemble approach to flight delay classification using data from Boston Logan International Airport, stacking consistently outperformed individual algorithms in terms of the evaluation metrics. This study, similar to mine, demonstrates stacking ability and robustness in complex aviation datasets.

In the context of this data research, where flights are influenced by numerous non-linear and interdependent factors such as weather, scheduling, and operational constraints, stacking ensemble methods provide us with a powerful solution. Stacking solutions are layered, and the flexible structure can better capture complex feature interactions and deliver more generalisable, accurate predictions. This method solves a complex problem in model selection, particularly in a high-pressure domain like aviation, where explainability and reliability are non-negotiable.

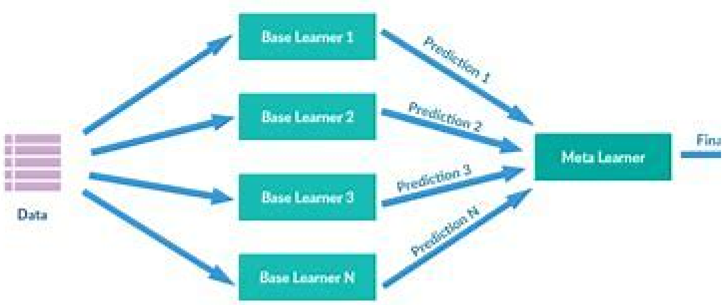


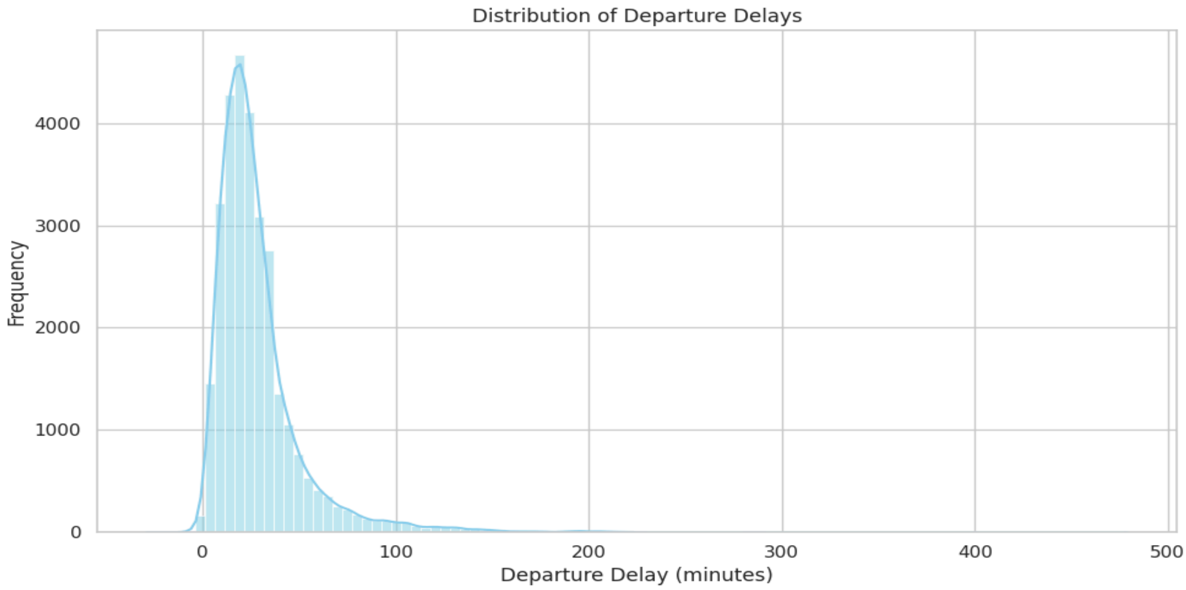
Figure 17 - Stacking Visualisation

CHAPTER 4 - EDA

(Exploratory Data Analysis)

4.1 EDA Introduction

This section outlines the exploratory data analysis (EDA) conducted on the combined dataset to investigate patterns and factors contributing to flight delays at Dublin Airport. The analysis focuses on temporal, operational, and environmental influences, with particular attention to



weather conditions and their economic and environmental impact, including CO_2 emissions.

It is crucial to start the analysis of the target variable, looking at the delays at Dublin airport during the specific period. The distribution of departure delays is highly right-skewed, as shown in FIG 18, with the majority of flights experiencing minimal delays and a smaller proportion affected by substantial delays exceeding 100 minutes. A boxplot of departure delays revealed several outliers beyond the typical range, which are likely to be due to exceptional operational disruptions as seen in FIG 19. Following this, a subset of negative delays was found and identified as data inconsistency and furthermore removed (n=11) to prevent disruption to the analysis.

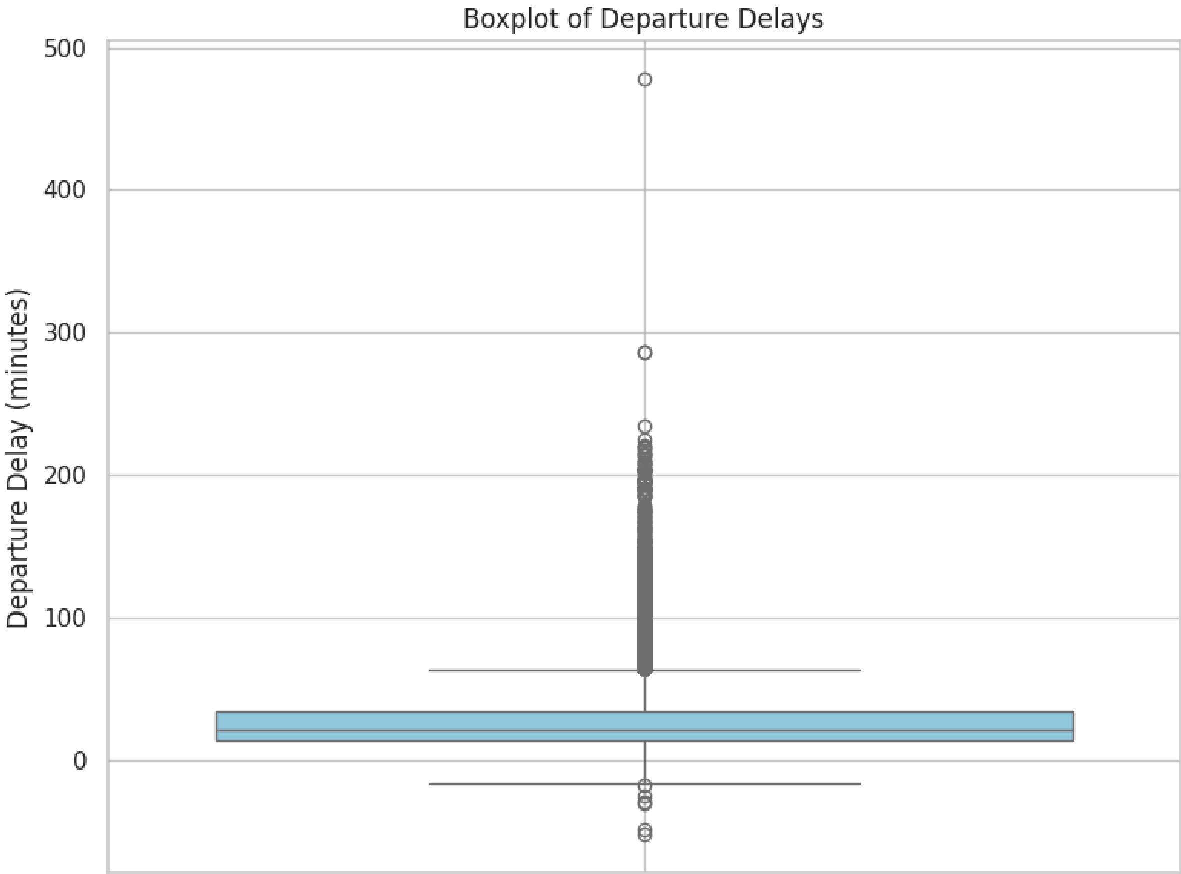


Figure 19 - Boxplot of Delay Distribution

The removal of the anomalies allowed for more effective interpretation of the delays as seen in FIG 20. Subsequent analysis revealed that the median departure delay was around 20–25 minutes, with the interquartile range indicating that most delays fell between 10 and 35 minutes. These findings suggest that while delays are generally short, operational disruptions can occasionally result in longer delays, which have a significant impact on emissions, as shown by the CO_2 emissions analysis in the later section 4.5.

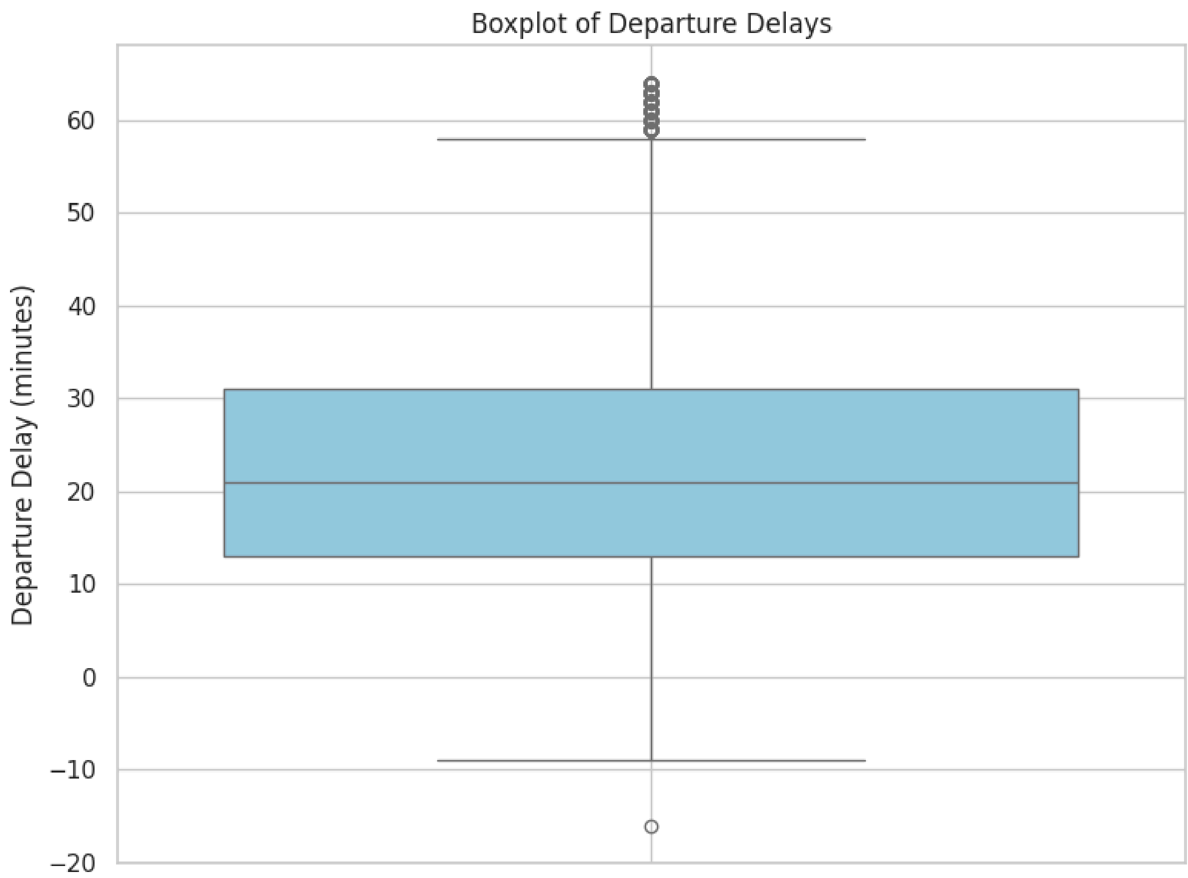


Figure 20 - Boxplot of Delays outliers removed

4.2 Temporal Patterns

This section aims to highlight the temporal aspects of flight delays at Dublin Airport. Specifically, it explores how delays can vary across different times of day, day of week and the months of the year. I aim to highlight this pattern as it can reveal underlying operational inefficiencies or periods of high congestion. This analysis can guide strategies for improving airport operations and resource allocation.

Delays by Day of the Week

The analysis of flight delays by day of the week reveals clear patterns in Dublin Airport's operations. As shown in Figure 21, Thursday and Friday experience the highest average delays, with delays exceeding 30 minutes on average. This aligns with the increased weekend traffic and the transition into the weekend, where higher flight volumes are likely to contribute to operational delays. Interestingly, Monday also shows relatively high delays, likely due to the return flights after the weekend. In contrast, weekdays like Tuesday and

Wednesday show more consistent delays at lower levels, indicating more efficient operational management during these mid-weekdays when traffic is generally lower.

An ANOVA test was conducted to assess whether there were significant differences in delays across the days of the week. The results indicated statistically significant differences in delays, as a p-value < 0.05, suggesting that flight delays vary depending on the day.

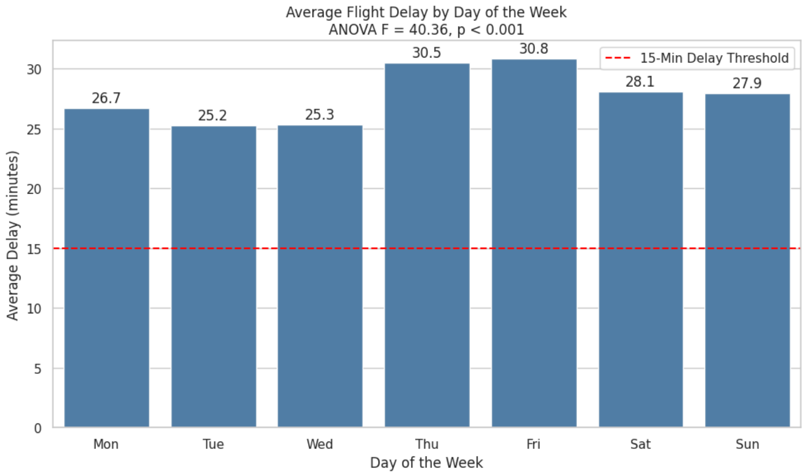


Figure 21 - Average flight delay by day of week

Delays by Hour of the Day

Clear analysis of flight delays by hour in FIG 22 reveals peak delay periods, particularly between 7:00–9:00 AM and 7:00–9:00 PM, with delays exceeding 30 minutes. This is likely due to higher flight volumes, airport congestion, and the accumulation of delays throughout the day. The morning peak aligns with Keenon (2013), which identified the busiest periods at European airports as 6:00 AM to 9:00 AM, reflecting increased operational load. In contrast, while late evening delays are notable, they are less pronounced. Early morning hours show significantly lower delays, likely due to reduced congestion and operational pressure. These findings highlight the importance of operational efficiency during peak times and the need for improved delay management during Dublin Airport's busiest periods.

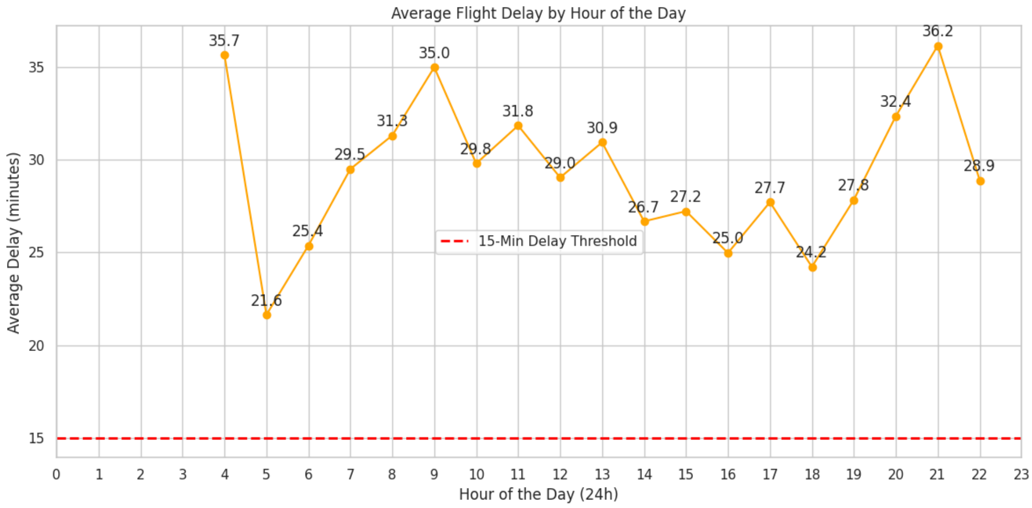


Figure 22 - Average flight delay by hour of day

Flight Volume vs Average delay by hour

The analysis of flight volume versus average delay by hour reveals clear patterns of congestion and operational inefficiencies throughout the day. As shown in FIG 23, the highest delays occur in the early morning and late evening hours, particularly between 7:00 AM and 9:00 AM and 7:00 PM and 9:00 PM, where delays exceed 30 minutes on average. The morning rush hour between 6:00 AM and 7:00 AM sees the highest flight volume, and delays begin to accumulate in the following hour, highlighting a knock-on effect as the airport struggles to manage the increased traffic. Similarly, the evening hours, particularly from 4:00 PM to 7:00 PM, show a similar delayed accumulation due to high traffic during these periods.

This EDA was essential to understanding the temporal dynamics of delays at Dublin Airport. Correlating flight volume and delays highlights the cumulative effects of operational bottlenecks during peak hours, both in the morning and evening. This provides valuable insights for targeting operational improvements during these high-traffic periods.

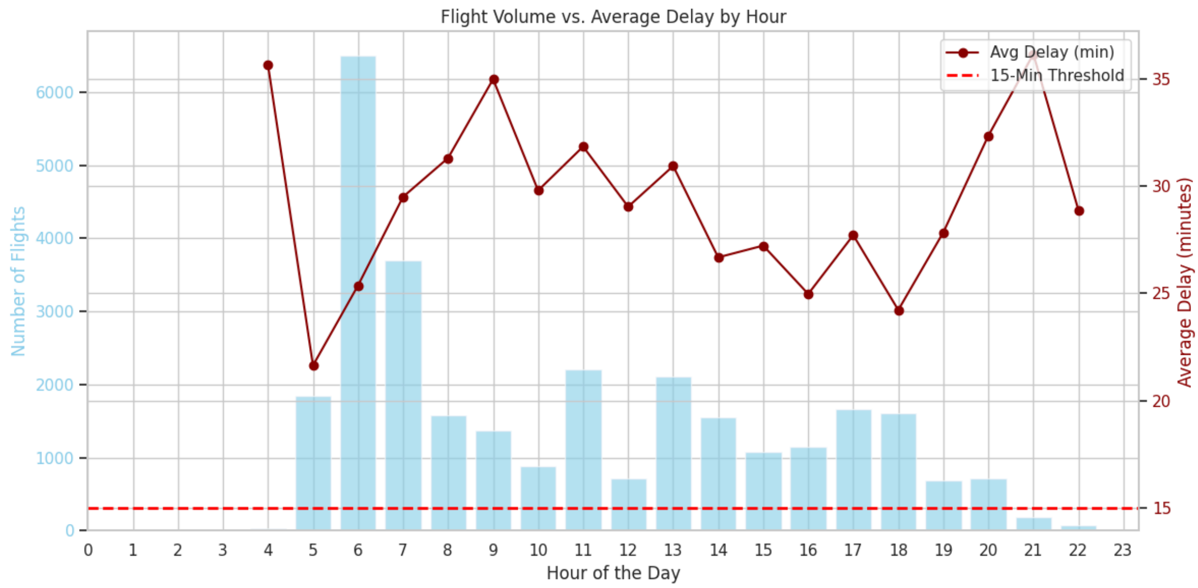


Figure 23 - Average flight delay by hour compared to volume

4.3 Operational Factors

Operational factors play a significant role in flight delays, as illustrated in FIG 24. The analysis reveals that the average flight delay increases as the flight duration rises. On top of this, we have discovered that short-haul flights have the lowest average delay of (11 minutes), whereas long-haul flights show the longest delays, reaching an average of nearly (39 minutes). This pattern is logical, as longer routes often involve extended boarding, more complex turnaround procedures, and a higher likelihood of cascading delays from earlier legs of multi-segment operations.

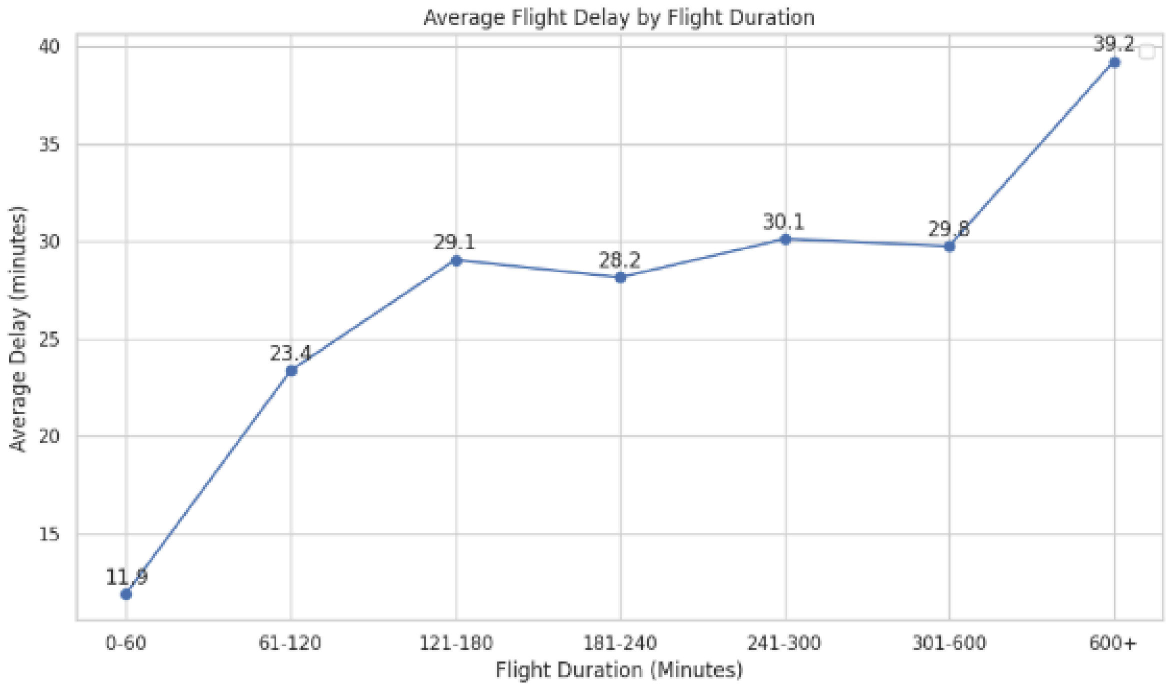


Figure 24 - Flight delay sorted by flight duration

An analysis of airline-specific delay rates reveals considerable variation in operational performance across carriers. As shown in FIG 25, Wizz Air, Virgin Atlantic, and Etihad Airways exhibit the highest proportions of delayed flights, with delay rates approaching 100%. However, this spike may reflect a limitation in the dataset, as the random selection of flights each day may not accurately represent overall airline performance. Notably, FIG 24 also highlights Irish carriers Aer Lingus and Ryanair in green to enhance comparative analysis. Both operate near or just below the average delay rate of 0.69. This suggests relatively stable performance at Dublin Airport. These insights are valuable for identifying potential operational inefficiencies and for benchmarking airline reliability at the airport.

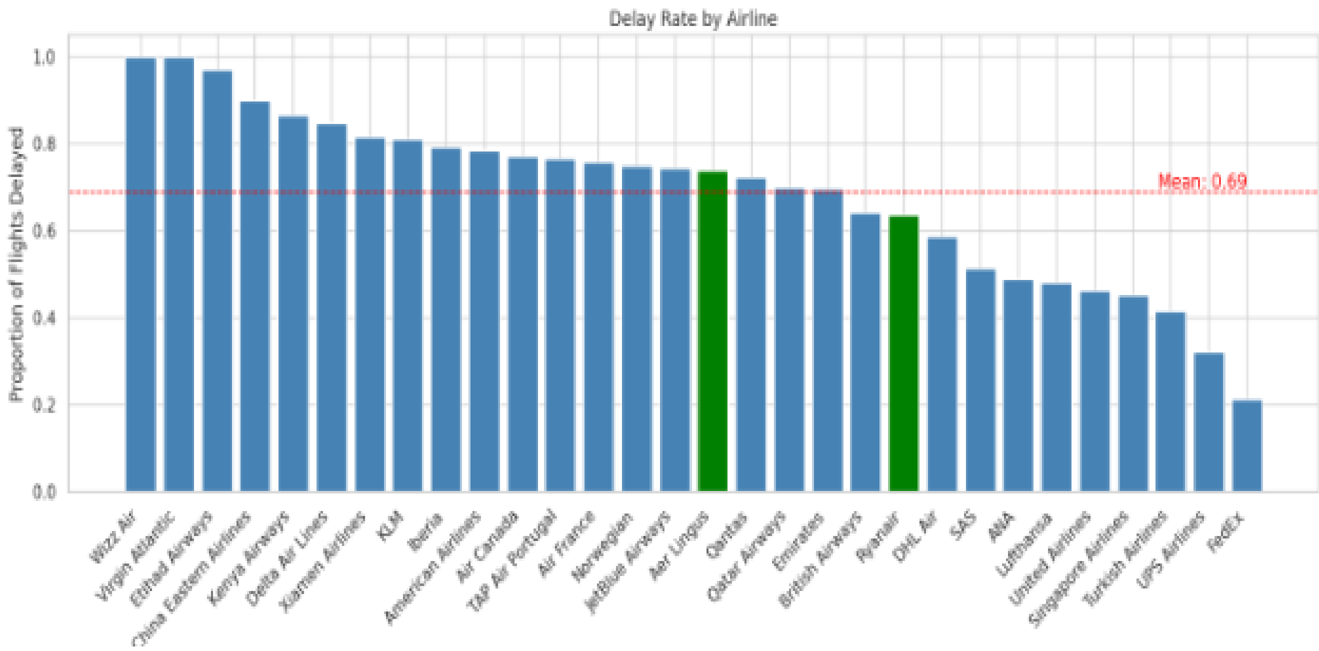


Figure 25 - Delay Rate by Airline

A common stereotype in the aviation industry is that low-cost airlines experience longer and more frequent delays. FIG 26 below highlights the average departure delays between low-cost and full-service airlines. Full-service carriers experience marginally longer delays on average of 28.8 mins compared to 27.7 minutes for low-cost airlines. While minimal the difference, this common perception of low-cost airlines being unreliable challenges that belief. This suggests that all airlines face similar operational pressures at Dublin Airport, emphasising the broader impact of external factors such as weather and congestion.

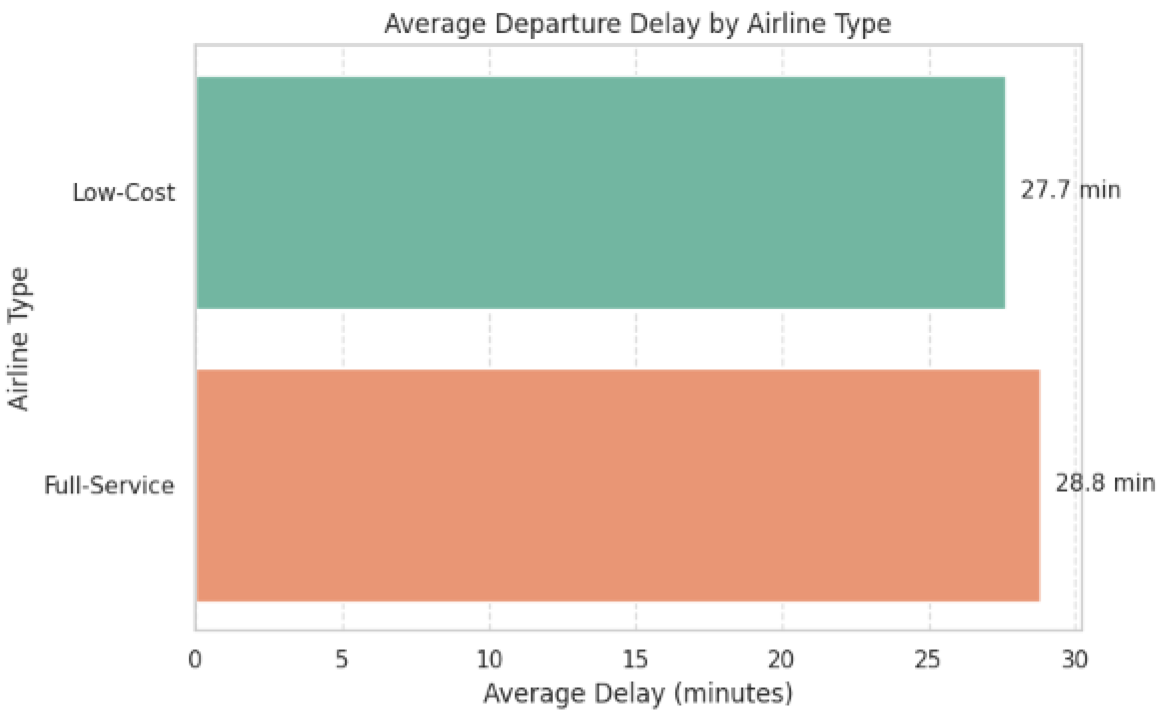


Figure 26 - Average Delay rate by airline type

4.4 Weather Impact

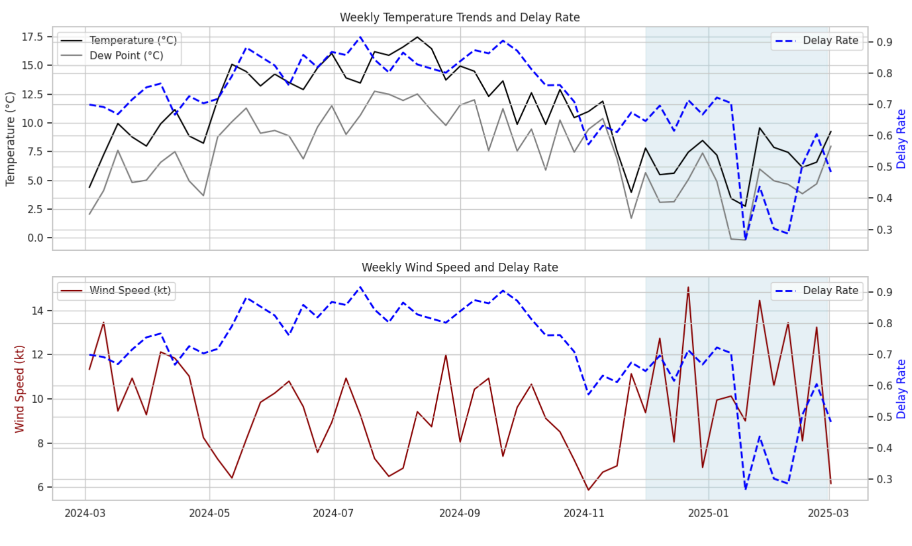


Figure 27 - Weekly temperature, wind speed, and delay rate trends over time

The exploratory data analysis presented in Figure 27 illustrates the complex relationships between weather conditions and flight delay rates at Dublin Airport. Contrary to initial expectations, the delay rate demonstrates a decreasing trend during the colder winter months (December–February), suggesting that harsh weather conditions do not directly correlate with increased delays. Similarly, FIG 27 (panel 2) shows that wind speed alone does not consistently align with elevated delay rates, indicating that wind is unlikely to serve as a robust standalone predictor of flight delays. These findings imply the influence of underlying operational factors, such as seasonal variations in flight scheduling or proactive delay mitigation strategies by airport management, which potentially offset adverse weather effects. Consequently, this analysis highlights the necessity of incorporating comprehensive operational and seasonal variables into the predictive modelling approach detailed in Chapter 5.

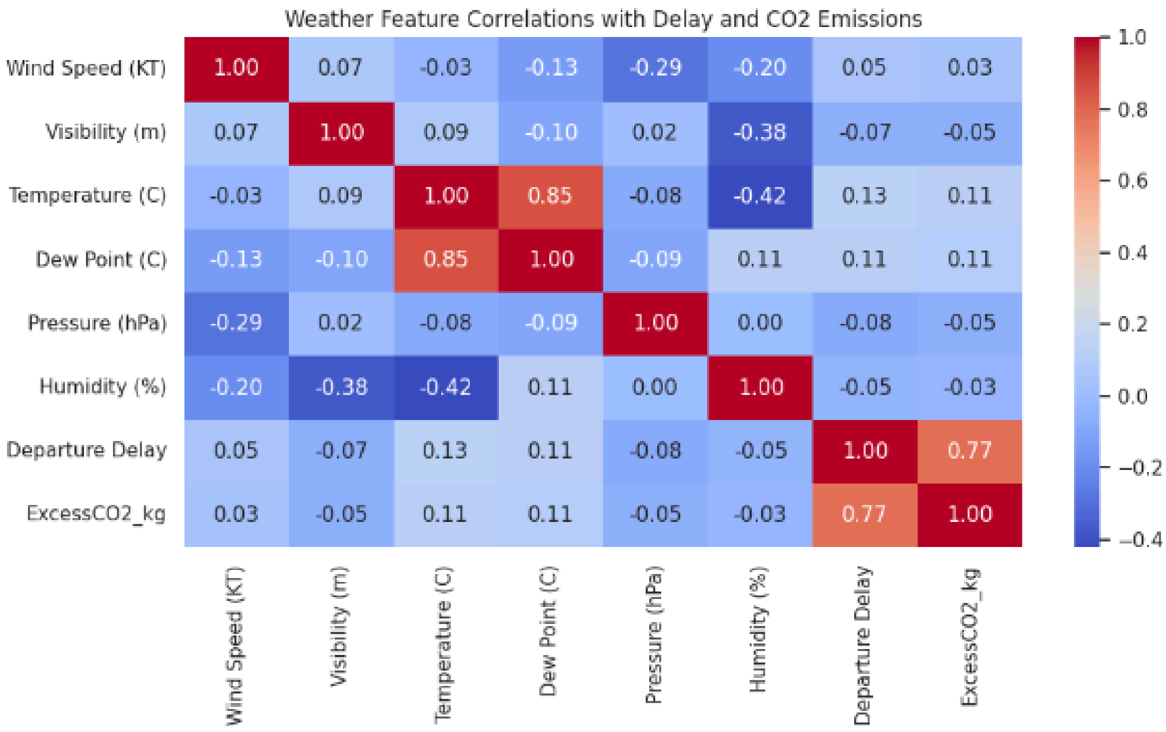


Figure 28 - Weather Variables Correlations

Further analysis in FIG 28 of the correlation between weather variables, departure delay, and excess CO_2 emissions reveals only weak individual associations between weather conditions and delays; temperature and humidity show the strongest correlations, though they remain modest. In contrast, a strong positive correlation (r = 0.77) is observed between departure delay and excess CO_2, which is expected, as delay duration directly contributes to emissions. These findings suggest that flight delays are influenced not by any single weather condition but by multivariate interactions, reinforcing the need for machine learning models that can capture complex, non-linear relationships.

	Variable	t-statistic	p-value	Significant (Y/N)
0	Wind Speed (KT)	2.07	0.0387	Yes
1	Visibility (m)	-4.41	0.0000	Yes
2	Temperature (C)	28.00	0.0000	Yes
3	Dew Point (C)	27.49	0.0000	Yes
4	Pressure (hPa)	-12.50	0.0000	Yes
5	Humidity (%)	-4.99	0.0000	Yes

Figure 29 - T-test on weather variables

Figure 29 above further explores the relationship between weather conditions and flight delays. The boxplots compare the distribution of key weather variables across delayed and non-delayed flights. These visualisations complement the t-test results presented in Fig 30, illustrating how weather factors such as temperature, humidity, and pressure differ between the two groups, even if the difference appears subtle. These results justify the inclusion of weather features in the machine learning models, where complex interactions and non-linear effects can be effectively captured.

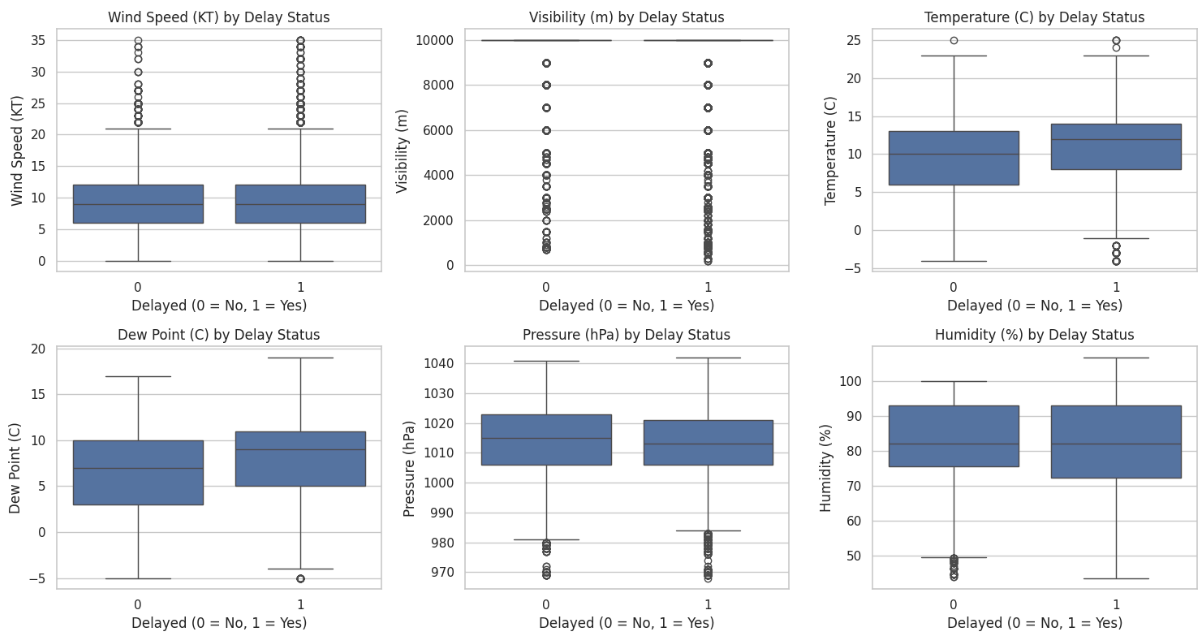


Figure 30 – Boxplots displaying the difference between weather conditions during both targeted variables

4.5 Environmental Impact

Sustainability is increasingly crucial, as regulations to combat climate change intensify across sectors. Dublin Airport is directly impacted by SDG 13 – Climate Action, which focuses on reducing greenhouse gas emissions. Flight delays at the airport contribute to excess CO_2 emissions, as aircraft continue to burn fuel while idling.

To estimate these emissions, aircraft were classified by size, and corresponding fuel burn rates from EUROCONTROL (2008) were applied (Section [2.4]). The excess CO_2 emissions per minute were calculated using the following formula:

$$\text{Excess CO}_2\ (\text{kg}) = \text{Delay (min)} \times \text{Fuel Burn Rate (kg/min)} \times 3.16$$

Figure 31 - Formula to calculate Excess Emission

The value 3.16 represents the CO_2 emissions produced per kg of fuel burned, a standard conversion factor in aviation.

Environmental Analysis

I wanted to highlight the daily excess CO2 emissions throughout the period under analysis. This revealed a downward trend across the year, as you can see in FIG 33. Despite a few periodic spikes, which are likely tied to peak travel days or unforeseen operational disruptions. This figure highlights how excess emissions often fluctuate day-to-day, but overall, this year has shown an improvement for Dublin Airport. This is a crucial direction for the future as regulations begin to tighten. To complement the analysis of daily and seasonal CO_2 fluctuations, emissions were further examined at the airline level.

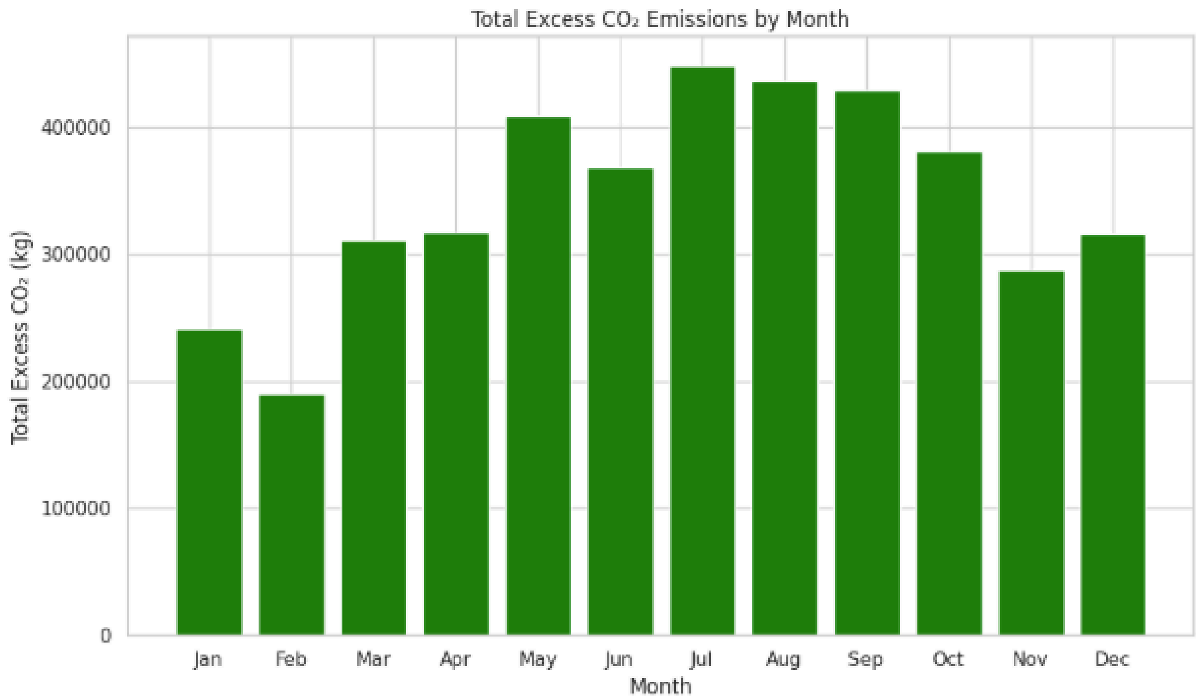


Figure 32 - Total Emissions by Month

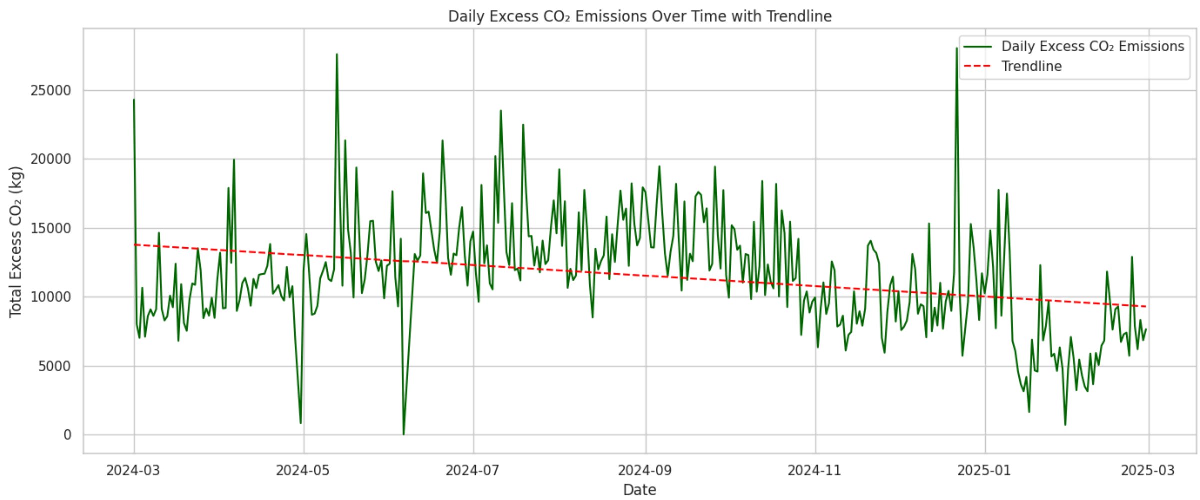


Figure 33 - Emissions over the analysed period

To assess the environmental performance of airlines operating out of Dublin Airport, FIG 34 and FIG 35 compare the top 10 emitters using two metrics: total excess CO_2 emissions and average emissions per flight. FIG 34 highlights major contributors such as American Airlines, Ryanair, and Aer Lingus, which reflect their high flight volumes. However, when normalised per flight in FIG 35, global carriers like Etihad Airways, United Airlines, and Delta Air Lines emerge as the least efficient. Notably, Etihad emits approximately 1.8 times more CO_2 per flight than Ryanair at Dublin Airport. This discrepancy is likely due to the operation of long-haul routes and the use of larger aircraft types. Assessing both total and per-flight emissions provides a more comprehensive understanding of airline-level environmental impact. These findings underscore the need for targeted mitigation strategies, particularly among international carriers that contribute disproportionately to the airport's overall emissions footprint.

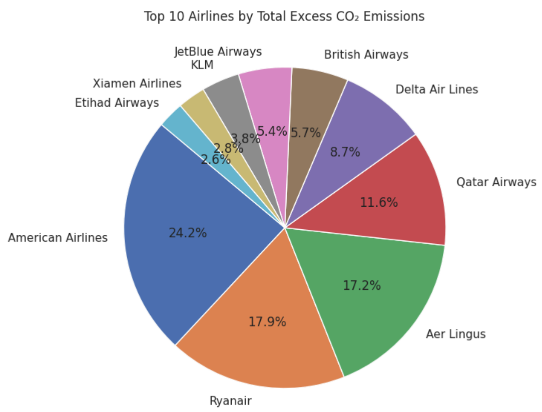


Figure 34 - Top 10 Airlines with Excess Emissions

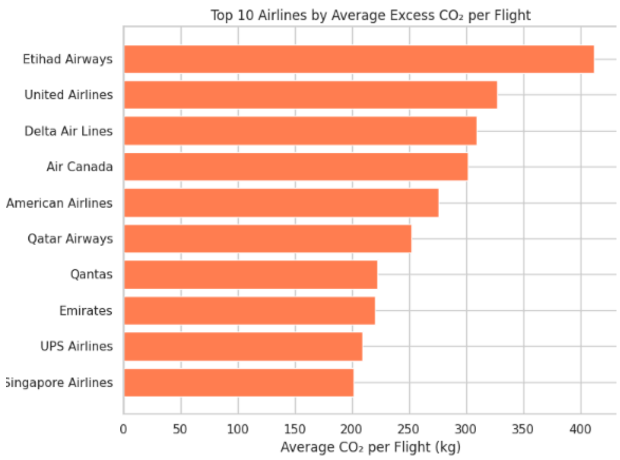


Figure 35 - Top 10 Airlines with the highest average excess emissions

FIG 36 illustrates the monthly relationship between average flight delays and total excess CO_2 emissions. The association between delays and emissions is further supported by a seasonal comparison of departure delays in FIG 37. This presents a distribution of delays during peak and off-peak months. The peak season exhibits a higher median delay and a wider distribution, with more frequent occurrences of delays exceeding 50 minutes. This pattern aligns with the monthly trend observed in FIG 36, where both average delays and emissions increase significantly during the summer months. The broader spread and elevated central tendency of delays during the peak period in FIG 37 suggest heightened operational inefficiencies, contributing to increased fuel burn and emissions. Together, these visuals provide strong evidence that extended delays during high-traffic periods play a significant role in the airport's overall environmental impact.

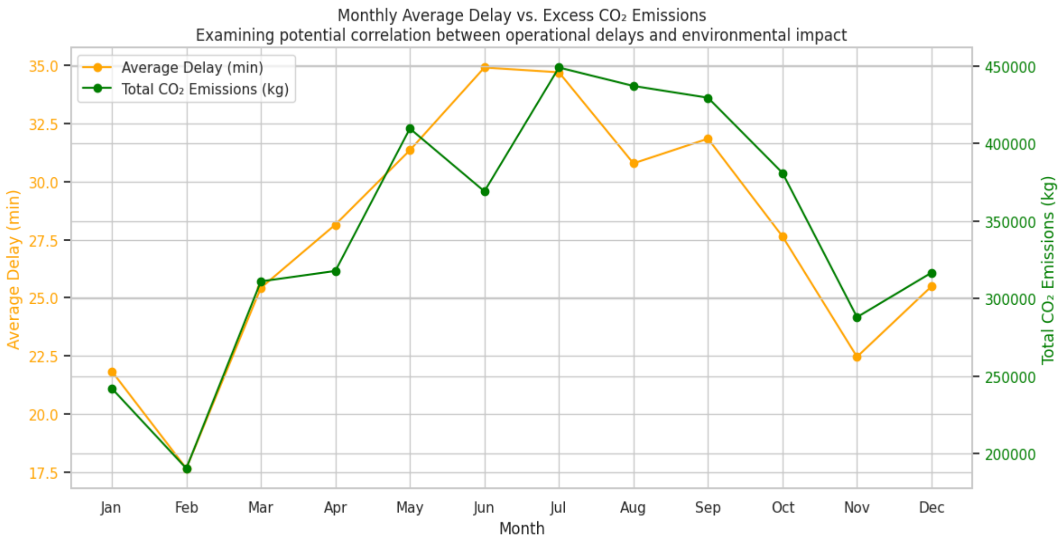


Figure 36 - Monthly average delay and CO_2 emissions showing correlation between delays and environmental impact.

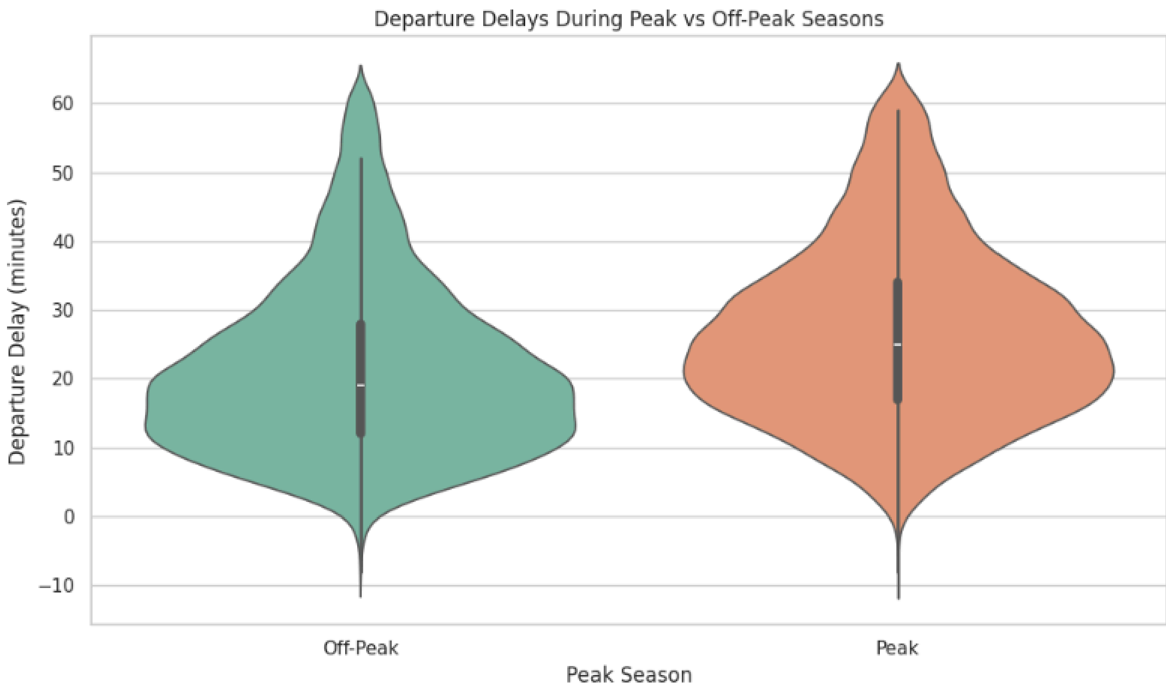


Figure 37 - Distribution of departure delays during peak and off-peak seasons

4.6 EDA Summary

This study's exploratory data analysis (EDA) revealed key insights into delay patterns and trends at Dublin Airport when looking at five various dimensions: delay distribution, temporal trends, operational factors, weather and environmental impact.

The distribution of delays was right-skewed, with most under 35 minutes and a select few exceeding 100 minutes. The median delay was 20-25 minutes when the outliers were removed to ensure accuracy. When investigating temporal analysis, peak delays were identified during 7-9am and 7-9pm. While the busiest days of the week often tended to be Thursdays, Fridays and Mondays. An ANOVA test confirmed statistically significant variation by the day of the week.

Operational insights suggested long-haul flights experienced the longest delays, averaging 39 minutes, while short-haul haul averaged up to 29.1 minutes at a high. A commonly held belief was challenged by the finding that low-cost carriers outperformed full-service airlines in terms of delayed performance.

A critical finding with regards to the weather variables found that individual factors had weak correlations with delays. This was further investigated when looking at multivariate relationships between weather variables in the machine learning application of this study. Environmentally, larger aircraft and peak periods contributed most to the excess emissions, as expected large and very large aircraft were less efficient on a per flight basis.

Overall, the EDA confirmed that delays arise from a mix of operational congestion, scheduling patterns and environmental scenarios, as a result this forms a critical foundation for predictive modelling in the following chapter.

Chapter 5 - Results and Findings

5.1 Results introduction

This chapter presents the findings of the machine learning models developed to predict flight delays at Dublin Airport, with a focus on weather-related features. The results are organised to reflect the progression of the analysis, beginning with baseline model evaluations, followed by the application of SMOTE to address class imbalance, and concluding with hyperparameter tuning to improve performance. Top-performing models were then combined using a stacking ensemble to further enhance predictive accuracy. In addition to standard evaluation metrics, the potential environmental and financial impacts of predicted delays were assessed to provide broader operational insight for Dublin Airport.

5.2 Experimental Setup

All experiments were conducted in Python 3.11 within the Google Colab environment. A range of open-source libraries were used throughout the project, including Scikit-learn for traditional machine learning models, XGBoost for gradient-boosted trees, TensorFlow for neural networks, imbalanced-learn for SMOTE application, and SHAP for model interpretability.

The dataset, consisting of historical flight operations and weather conditions, was cleaned and normalised using StandardScaler. An 80/20 train-test split was applied to evaluate out-of-sample performance. Initial baseline models were trained on the imbalanced dataset, after which SMOTE was applied exclusively to the training data to balance class distribution before retraining. Various supervised classification models were selected for comparison, as seen in Section 5.3. All the models were evaluated using standardised classification metrics.

To reduce the risk of overfitting and ensure generalisability, several techniques were applied. These included hyperparameter tuning via GridSearchCV, dropout, and early stopping in neural network models, and the use of ensemble methods such as Random Forest and Stacking to enhance robustness. The path can be seen below in Fig 38.

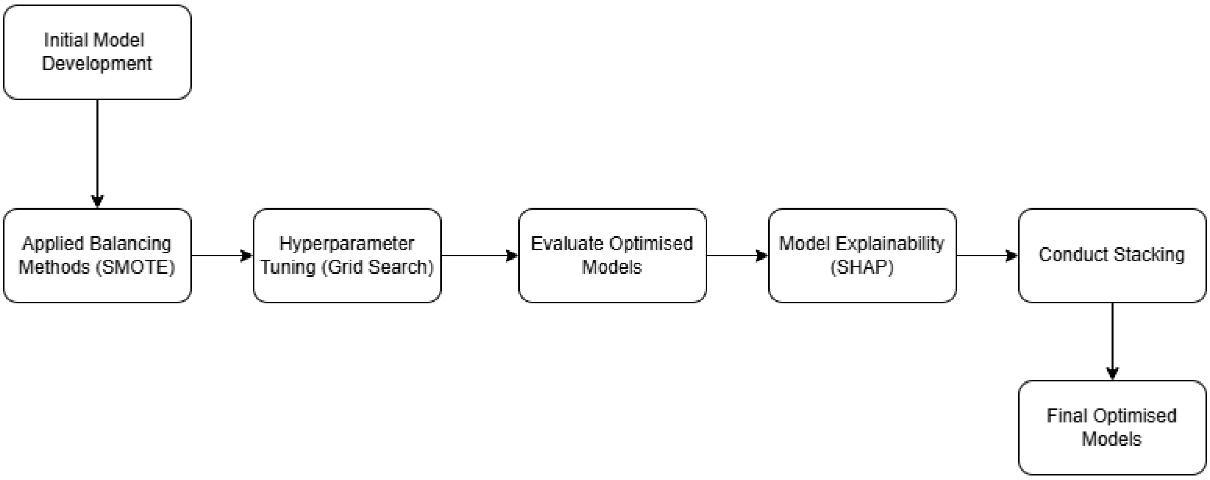


Figure 38 - Workflow of model development, optimisation, explainability, for flight delay prediction

Financial Analysis Set up

The development of financial analysis was a critical component in ensuring the real-world relevance and overall success of this study. A scaling methodology was employed to estimate financial delay costs based on aircraft size. Rather than applying a uniform cost across all aircraft types, which would not accurately reflect operational realities, this study utilised industry data from EUROCONTROL to assign size-specific cost estimates seen in Table 1. (https://www.eurocontrol.int/sites/default/files/content/documents/single-sky/pru/publications/other/cost-of-air-transport-delay-in-eu-ita.pdf)

Table 1 - Financial Delay Values

Aircraft Size	Gate Delay Cost (€)	Taxi Delay Cost (€)	Average Delay Cost (€)
Small	€72.00	€99.00	€85.50
Medium	€102.00	€134.00	€118.00
Large	€161.00	€196.00	€178.50
Very Large	€188.00	€229.00	€208.50

Specifically, average delay costs were calculated as seen above in the Table for the respective aircraft size. This approach addresses the inherent variation in operating costs across aircraft categories and deals with aircraft sensitive delay costs across aircraft categories throughout the dataset. As the type of delay was not collected in the dataset, consequently, the average operational delay cost was applied based on aircraft size. Although this introduces some limitations to the precision of the financial model, the use of industry standard benchmarks provides a realistic and scalable framework for estimating the broader financial impact of delays.

5.3 Hyperparameter Tuning and Parameter Decisions

Hyperparameter tuning is essential for enhancing the performance and generalisability of machine learning models. Unlike parameters learned during training, these hyperparameters are set beforehand and influence the model's ability to predict unseen data. This study implemented GridSearchCV with three-fold cross validation to systematically identify the optimal hyperparameter combinations which can be seen below. As mentioned above, the ROC AUC metric is used in this research to evaluate the model's performance, given its ability to handle class imbalance.

To further address the imbalance in the dataset, Synthetic Minority Oversampling Technique, also known as (SMOTE) was applied to the training data, ensuring the model could better capture patterns in the minority (on-time) class. This combination has been shown to significantly improve model performance in similar studies, Sujay, (2024)

Sujay (2024) highlights notable gains after integrating SMOTE and hyperparameter optimisation across models such as Random Forest and SVM, leading to metrics such as accuracy, precision and recall. Balasubramanian et al. (2017) reinforced these findings, demonstrating robustness in models exposed to imbalance and noisy aviation datasets.

Table 2 - XGBoost optimal parameters

XG Boost	
n_estimators	200
max_depth	7
learning_rate	0.2
colsample_bytree	1
subsample	1

Table 3 - Gradient Boosting Methods optimal parameters

Gradient Boosting Methods	
learning_rate	0.2
max_depth	5
n_estimators	200

Table 4 - Light Gradient Boosting Methods optimal parameters

Light Gradient Boosting Methods	
learning _rate	0.1
max_depth	7
n_estimators	200
num_leaves	50

Table 5 - Logistic Regression optimal parameters

Logistic Regression	
learning_rate	1
class _weight	balanced
penalty	L2
solver	saga

Table 6 - Random Forest optimal parameters

Random Forest	
max_depth	None
max_features	sqrt
min_samples_leaf	1
min_samples_split	2
n_estimators	200

Table 7 - KNN optimal parameters

KNN	
knn_metric	manhattan
knn_n_neighbours	3
knn_weights	distance

Table 8 - ADA optimal parameters

ADA Classifier	
learning_rate	1
n_estimators	200

Table 9 - MLP (neural network) optimal parameters

Neural Networks (MLP)	
early stopping	TRUE
learning rate	0.001
hidden layers	(128, 64)
alpha	0.001
solver	adam
max iterations	1000
random state	42

5.4 Model Performance Comparison

Below in Table 10 a comprehensive evaluation of various machine learning models was applied to flight delay prediction at Dublin Airport. Models are ranked from best to worst based on accuracy, with ties resolved using the AUC (Area Under the Curve) score. The results demonstrate that the Stacking ensemble model is the superior model, which combines XGBoost, Random Forest, and a Multi-Layer Perceptron (MLP), a form of Neural Network, using logistic regression as the meta-classifier. The Stacking model achieved an accuracy of 86% and an AUC score of 0.91, indicating a strong ability to predict between delayed and non-delayed flights 86% of the time. Its performance suggests that combining diverse model types (tree-based, neural, and regression) captures the complex patterns in flight delay data, particularly when influenced by multiple weather and operational variables.

Table 10 - ML optimised results

Ranked Classification Models by Accuracy and AUC Score

	Model	Model Accuracy	Precision	Recall	F1 Score	AUC SCORE
0	Stacking 1 (XGB + RF + MLP + LG)	0.86	0.85	0.86	0.85	0.91
1	Random Forest	0.85	0.85	0.85	0.85	0.91
2	XGBoost	0.83	0.83	0.83	0.83	0.87
3	Light Gradient Boosting	0.81	0.80	0.81	0.80	0.85
4	Gradient Boosting	0.80	0.80	0.80	0.80	0.84
5	SVM	0.76	0.75	0.76	0.72	0.74
6	KNN	0.74	0.76	0.74	0.75	0.77
7	MLP (Neural Networks)	0.72	0.76	0.72	0.73	0.79
8	Convolutional Neural Networks	0.69	0.73	0.69	0.70	0.75
9	Ada Classifier	0.69	0.72	0.69	0.70	0.73
10	Logistic Regression	0.64	0.70	0.64	0.65	0.68
11	Naive Bayes	0.61	0.69	0.61	0.63	0.66

One key advantage of the Stacking model was its strong performance on the imbalanced dataset, where on-time flights were represented by class 0. As shown in Table 11, the F1

*Table SEQ Table * ARABIC 11 - Stacking Results*

```
Accuracy: 0.8563
Classification Report:
              precision    recall  f1-score   support

           0       0.76      0.69      0.73      1650
           1       0.89      0.92      0.90      4328

    accuracy                           0.86      5978
   macro avg       0.83      0.81      0.81      5978
weighted avg       0.85      0.86      0.85      5978

ROC AUC Score: 0.9126
```

score reflects the model's ability to balance precision and recall for each class, and the Stacking model displayed the most balanced performance among all models tested. Notably, it showed the strongest ability to predict the minority class (on-time flights), outperforming the other models in this metric. Its performance suggests that combining diverse model types (tree-based, neural, and regression) captures the complex patterns in flight delay data, particularly when influenced by multiple weather and operational variables.

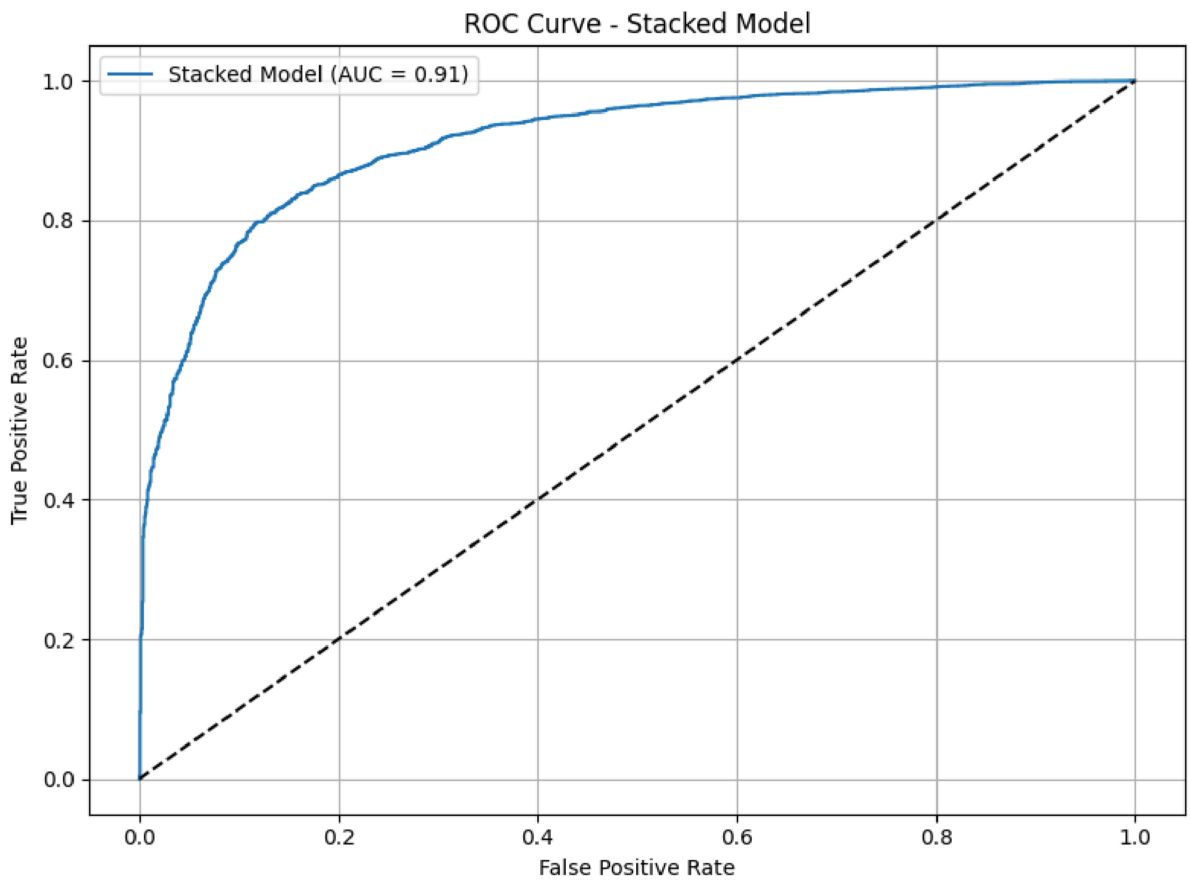


Figure 39 - Stacking ROC graph

The superior performance of the Stacking model (AUC = 0.91) seen in FIG 39, underscores its potential for real-time deployment at Dublin Airport, where accurate delay prediction can inform operational decisions such as gate assignments, runway scheduling, and early delay mitigation strategies. This model excels due to its ability to leverage the strengths of diverse algorithms, resulting in enhanced robustness and balanced performance, particularly for complex classification tasks such as flight delay prediction. Its ensemble nature allows it to incorporate deep learning, tree-based methods and regression, capturing a wide range of patterns and relationships effectively.

Closely following the stacking model was the Random Forest classifier, which achieved an accuracy of 85% and an identical AUC score of 0.91, highlighting its strong performance in modelling nonlinear relationships, as discussed in Section 3.6.2. XGBoost also performed well, with consistently high scores across all evaluation metrics, including an AUC of 0.87, reflecting its excellent ability to highlight differences between classes.

A clear pattern emerges when comparing the overall performance across all models: ensemble methods such as Stacking, Random Forest, Gradient Boosting, and LightGBM consistently outperformed simpler, standalone models. These advanced models demonstrated higher predictive accuracy and reliability, whereas traditional algorithms such as Logistic Regression and Naive Bayes showed the weakest performance, with accuracy scores of 0.64 and 0.61, respectively. The variation in model performance accuracy is visible in FIG 40. These findings strongly suggest that flight delay classification tasks benefit significantly from more sophisticated, ensemble-based approaches that can better accommodate the complex and multifaceted nature of aviation-related data.

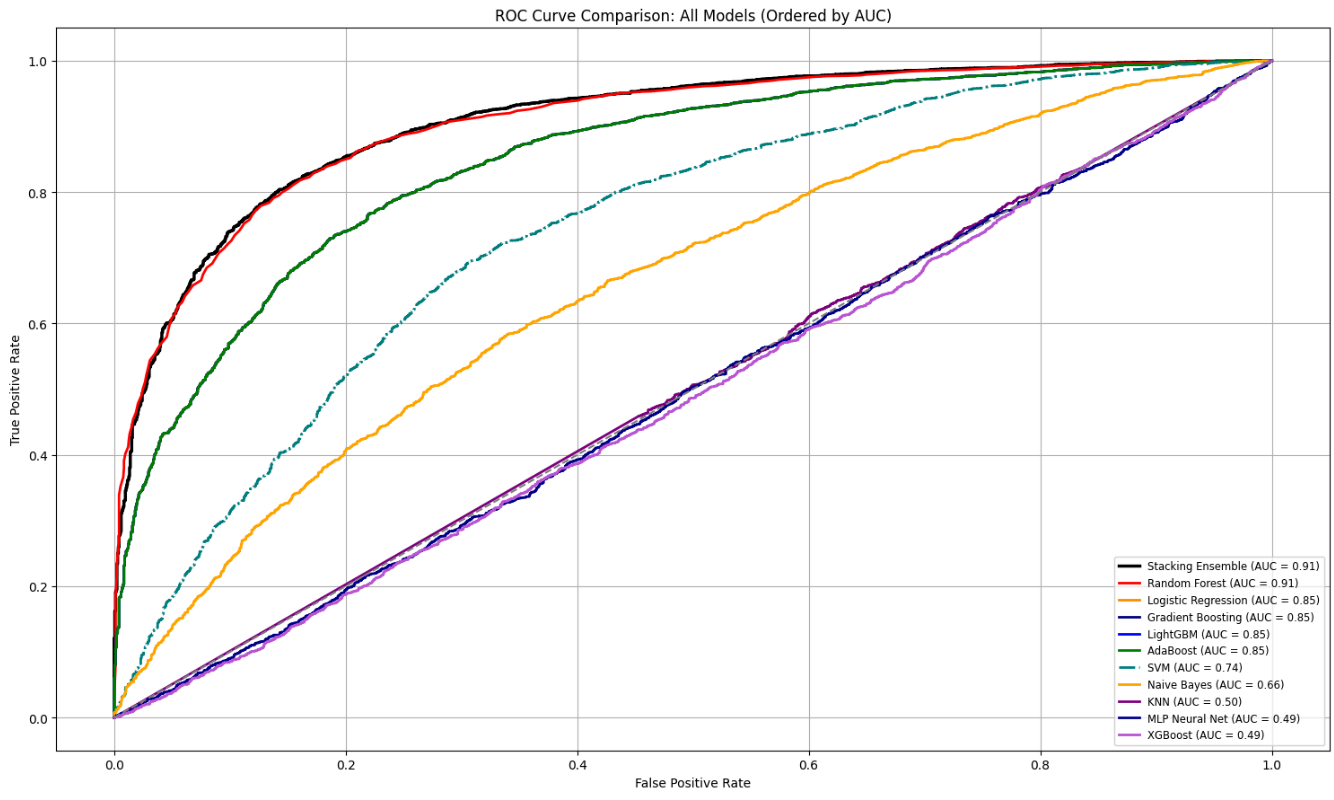


Figure 40 - Comparison of ROC graphs

5.5 Qualitative Results and Explainability

To complement model performance metrics and enhance interpretability, particularly in contrast to previous studies, SHAP (Shapley Additive Explanations) was applied in this research to provide both global insights into feature importance. SHAP enables a transparent understanding of which features most strongly influence the model's overall predictions, while also offering detailed explanations of how individual predictions are formed for local interpretation. This level of interpretability is particularly important for flight delay prediction, where operational decisions must be justified and explained. SHAP was therefore used not only to validate model behaviour but also to support the practical deployment of predictive models, offering actionable insights for stakeholders such as air traffic control and airport operations. These stakeholders require not only accurate forecasts, but also clear reasoning behind predictions to effectively implement mitigation strategies.

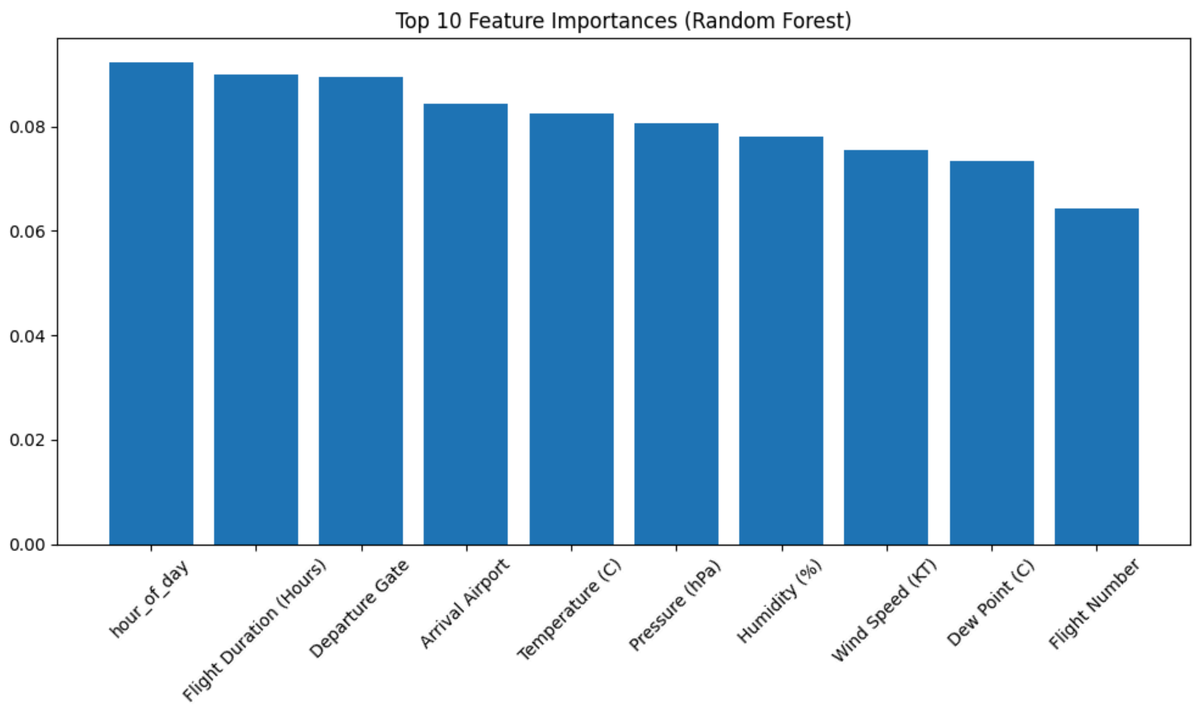


Figure 41 - Random Forest feature importance

As shown in FIG 41, the most influential features in the Random Forest model include the time of day, flight duration and departure gate. This suggests that operational factors are significant determinants of the delay likelihood, reflecting peak congestion periods. These features align with the findings revealed in the XGBoost model, reinforcing their consistent predictive value across different models. Several meteorological variables, such as Temperature, Wind Speed and Humidity, are also among the top contributors to classification. This pattern is evident in other strong-performing models, reinforcing that both operational efficiency and weather dynamics play a central role in delay prediction.

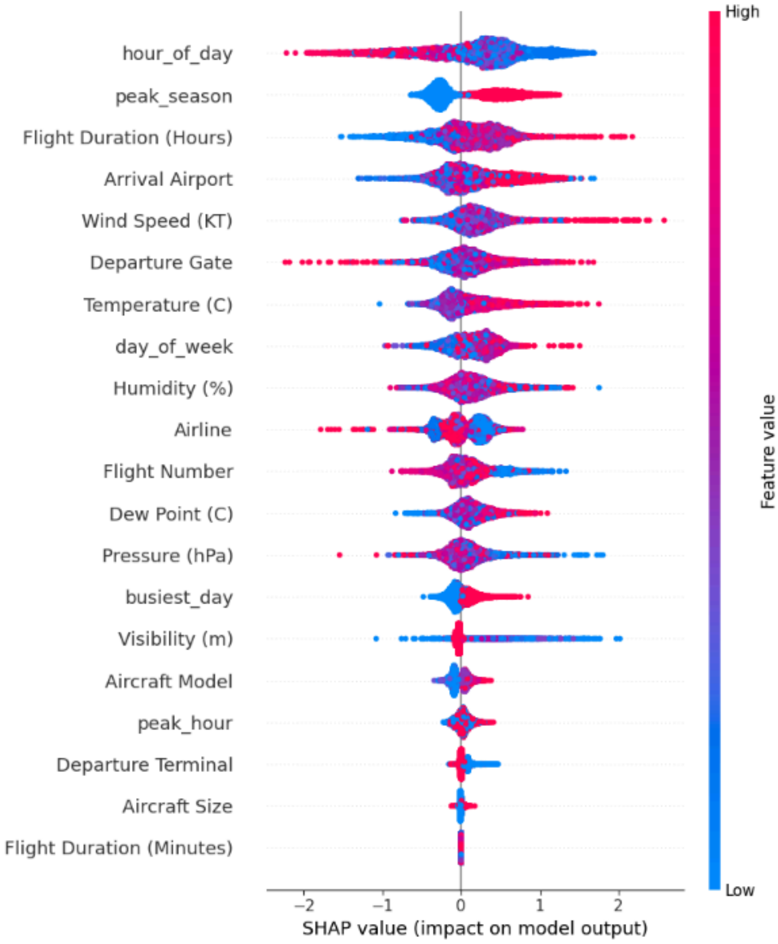


Figure 42 - SHAP summary plot showing feature importance and value impact on delay predictions.

The SHAP summary plot seen in Fig 42 for the XGBoost model further supports the findings by highlighting the global impact. It is evident that temporal and operational features such as the hour of day, flights during peak season and flight duration have the most influential impact on predictions, indicating that the model consistently relies on these variables for classification. Focusing on the most influential meteorological features, wind speed, visibility and temperature also rank highly, emphasising the role of weather in the delay dynamics specific to Dublin Airport.

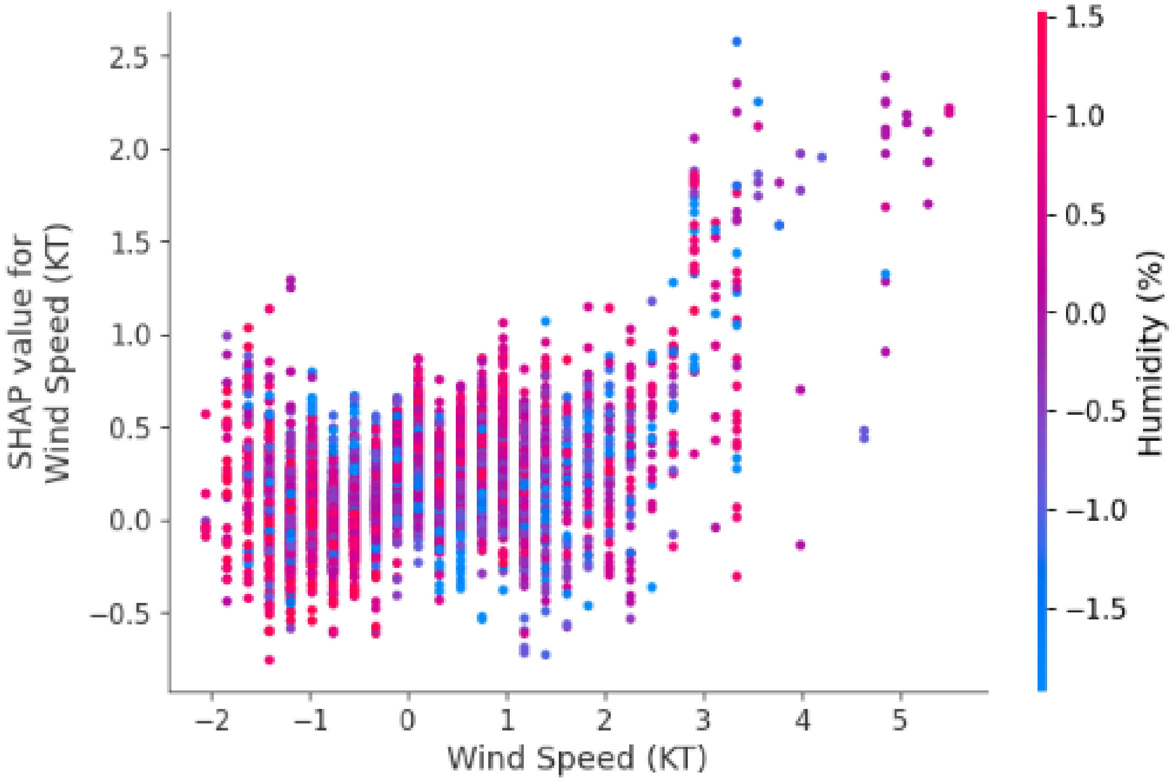


Figure 43 - SHAP plot showing wind speed's impact on delay prediction, coloured by humidity.

Focusing on the SHAP dependence plot for Wind Speed, which emerged as the most influential meteorological feature, FIG 43 reveals a clear upward trend in SHAP values as wind speed increases. This indicates a strong relationship between higher wind speeds and the likelihood of flight delays. When investigating the interaction with Humidity, the plot suggests that humid and extremely windy conditions have a compounding effect on delay predictions.

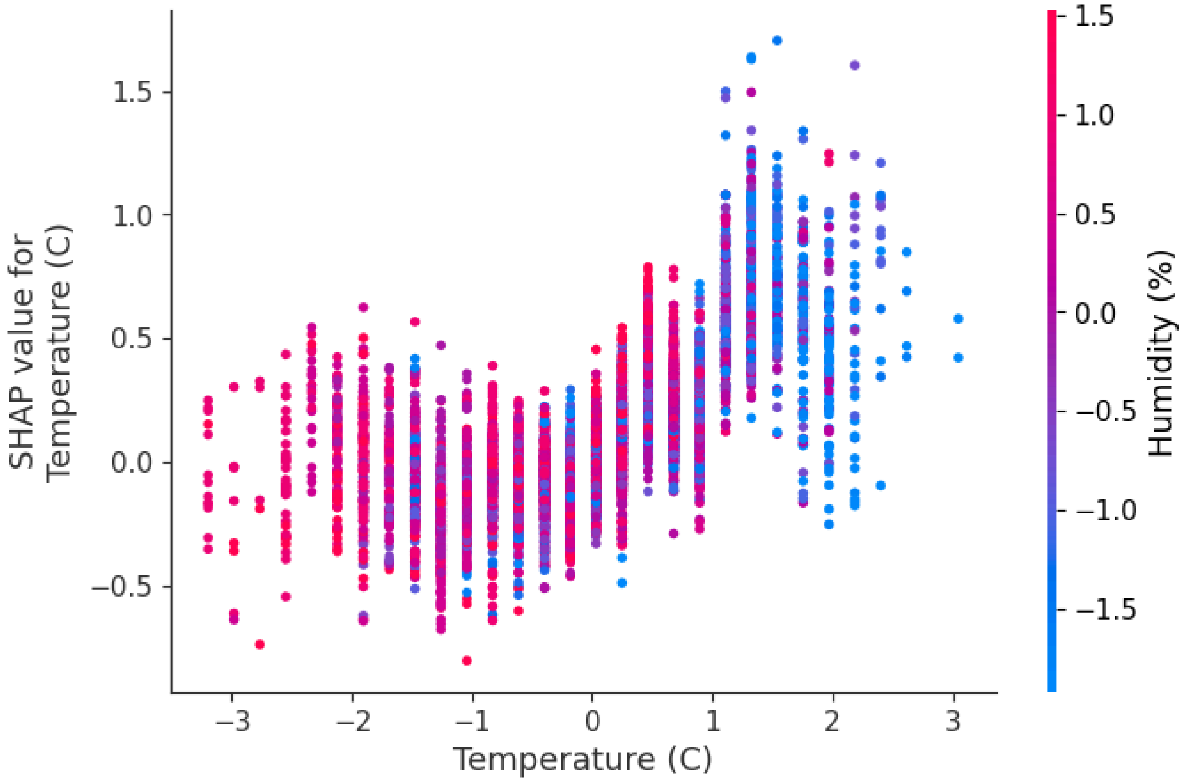


Figure 44 - SHAP plot showing temperature's impact on delay prediction, coloured by humidity.

A similar interaction is evident in FIG 44, where higher temperatures are associated with increased contribution to predicting delay risk, particularly when paired with high humidity. This supports findings from the EDA concluding that weather impacts on delays arise from interacting conditions, such as temperature, wind and humidity.

This finding directly contrasts with the results of Hatıpoğlu and Tosun (2024), who reported that weather variables had little to no impact on flight delays in the Turkish context. Such a discrepancy may be attributed to differing regional weather patterns and operational conditions, further underscoring the importance of location-specific research in flight delay modelling.

Since ensemble methods are the highest performers and most utilised methods in this domain, it was decided to investigate how each model in the stacking algorithm affected the decision. One limitation of SHAP is that it cannot be applied to a full stack, so the meta-learner (logistic regression) was analysed instead. SHAP values were computed using the output probabilities from the base learners (Random Forest, XGBoost, and MLP) as input features. Below in FIG 45, it displays how the meta-learner relied heavily on the Random Forest model, with minimal impact from MLP and XGBoost as base learners. Concluding the final decision within a stacking model is largely driven by the most robust base learner from the original testing.

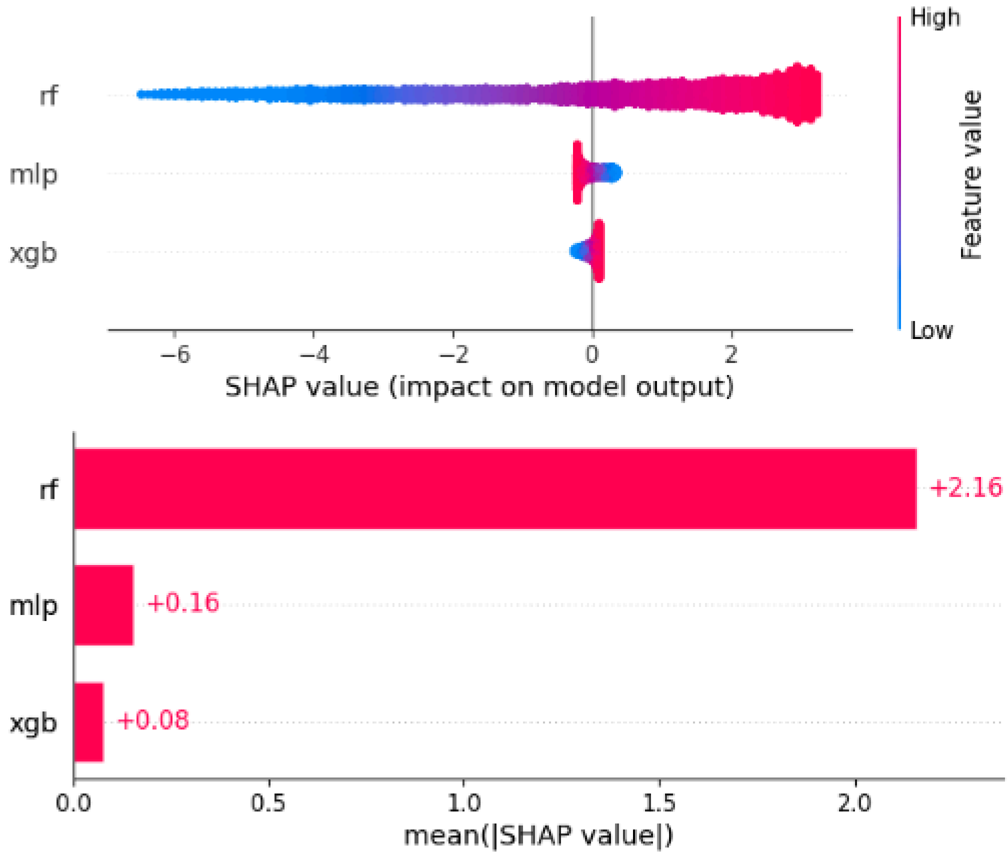


Figure 45 - SHAP values showing each base model's contribution to the stacking ensemble output.

5.6 Review of Research Question

This study confirms that machine learning can accurately predict flight delays at Dublin Airport, at the same time providing meaningful insights into their financial and environmental impacts. The stacking ensemble proved this with its impressive performance, especially compared to previous studies in the domain, achieving (86% accuracy, AUC 0.91). SHAP analysis revealed key predictors such as departure time, flight distance and weather conditions. The analysis quantified the delay-related emissions and costs, showing long haul flights and peak seasons to be major contributors. These results demonstrate that machine learning can offer both predictive insights and practical value.

Chapter 6 – Discussion

6.1 – Research Impact

6.1.1 - Environmental impact

This study found that flight delays at Dublin Airport contribute significantly to excess CO_2 emissions and place additional environmental strain on top of existing operational disruptions. As demonstrated in the environmental exploratory data analysis, a delay-based emissions estimation formula, accounting for both aircraft size and fuel burn rates, was used to quantify excess CO_2 emissions for all delayed flights in the dataset. The results revealed that emissions were highest during peak traffic windows (07:00–09:00 and 16:00–19:00) and during the summer travel season. These periods coincided with the longest median delay durations, thereby amplifying their environmental impact.

Aircraft size was identified as a significant contributor to emissions during delays, particularly for long-haul operations. As illustrated in FIG 46, large and very large aircraft produced the highest levels of excess CO_2, highlighting the compounded environmental cost of long-haul flight disruptions compared to shorter regional services. These findings underscore the importance of incorporating aircraft type and route profile into any future emissions mitigation strategies at Dublin Airport.

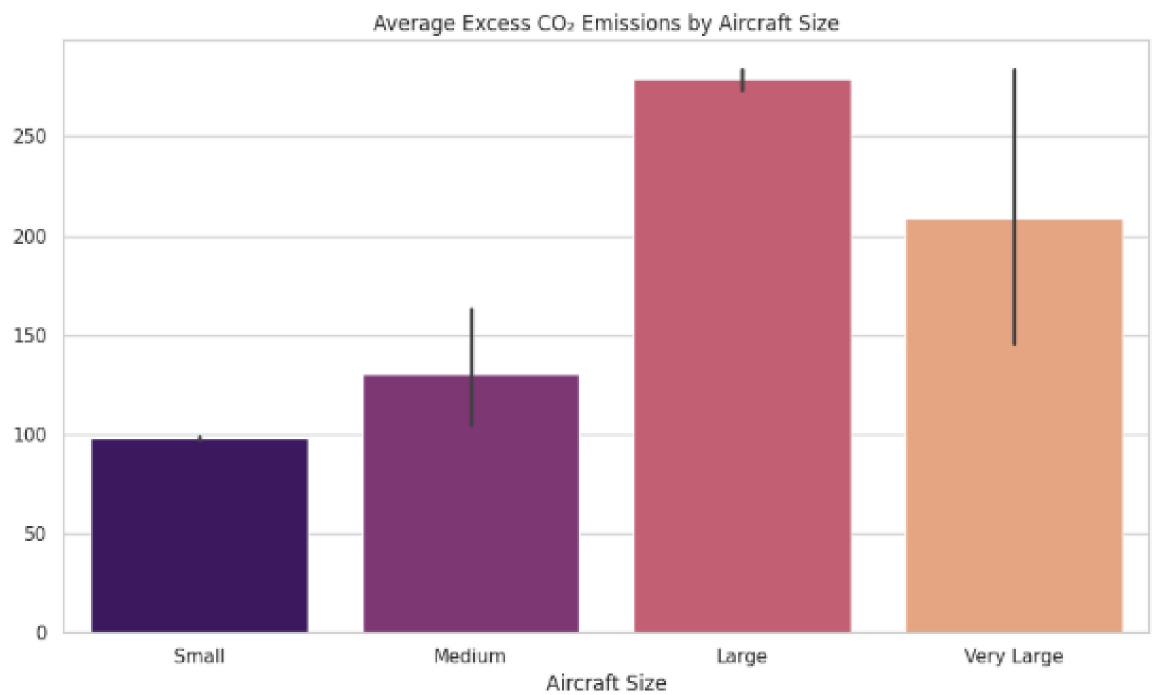


Figure 46 - Average excess CO_2 emissions by aircraft size category

When emissions were analysed by airlines, American Airlines, Ryanair, and Aer Lingus emerged as the top contributors to total excess CO_2 emissions, primarily due to their high flight volumes. However, when emissions were normalised on a per-flight basis, Etihad Airways, United Airlines, and Delta Air Lines were found to be the least fuel-efficient. This is largely attributable to the use of long-haul aircraft, which consume significantly more fuel and produce substantially higher emissions when delayed.

Looking to the future, these insights are particularly relevant as Dublin Airport works toward its sustainability commitments, including the goal of achieving net-zero emissions by 2050. These are ambitious targets with considerable regulatory and reputational implications if not met. The integration of predictive modelling, as developed in this research, could support real-time delay prevention and smarter operational decision-making, ultimately helping to reduce unnecessary fuel consumption and emissions.

Ultimately, this study not only identifies the key contributors and drivers of delay-related emissions but also presents a clear, data-driven path forward. To remain resilient and environmentally responsible, Dublin Airport must embed predictive analytics into its operational strategy and continue to prioritise both efficiency and sustainability in its daily decision-making.

6.1.2 - Financial Impact

The financial analysis conducted in this study provides a deeper understanding of the economic impact of departure delays at both the airline and industry levels. As outlined in Section 5.1, a cost-scaling methodology was implemented based on aircraft size to move beyond a flat-rate cost model and produce more realistic, aircraft-sensitive delay cost estimates.

Figure 47 supports the finding that airlines operating larger aircraft, particularly long-haul carriers such as Etihad Airways and United Airlines, incurred the highest average delay costs. This reflects not only higher crew and fuel costs but also increased passenger dissatisfaction and greater operational risk.

Further insight is shown in Figure 48, where a strong correlation was observed between financial delay cost and excess CO_2 emissions. Airlines with higher financial losses per delay also tended to emit more CO_2, underlining the dual burden of delays, both economic and environmental, and highlighting inefficiencies within operations.

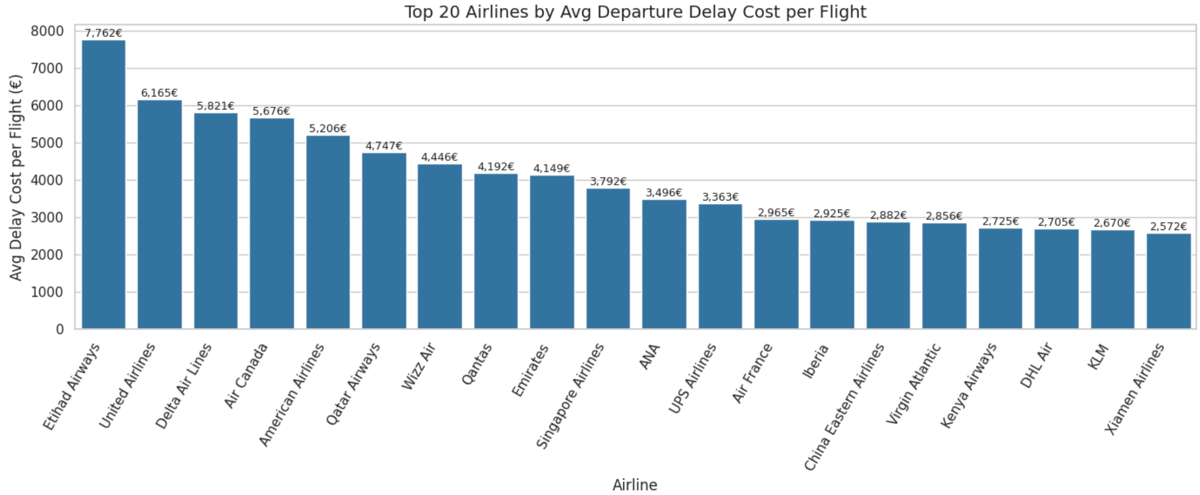


Figure 47 - Top 20 airlines ranked by average departure delay cost per flight

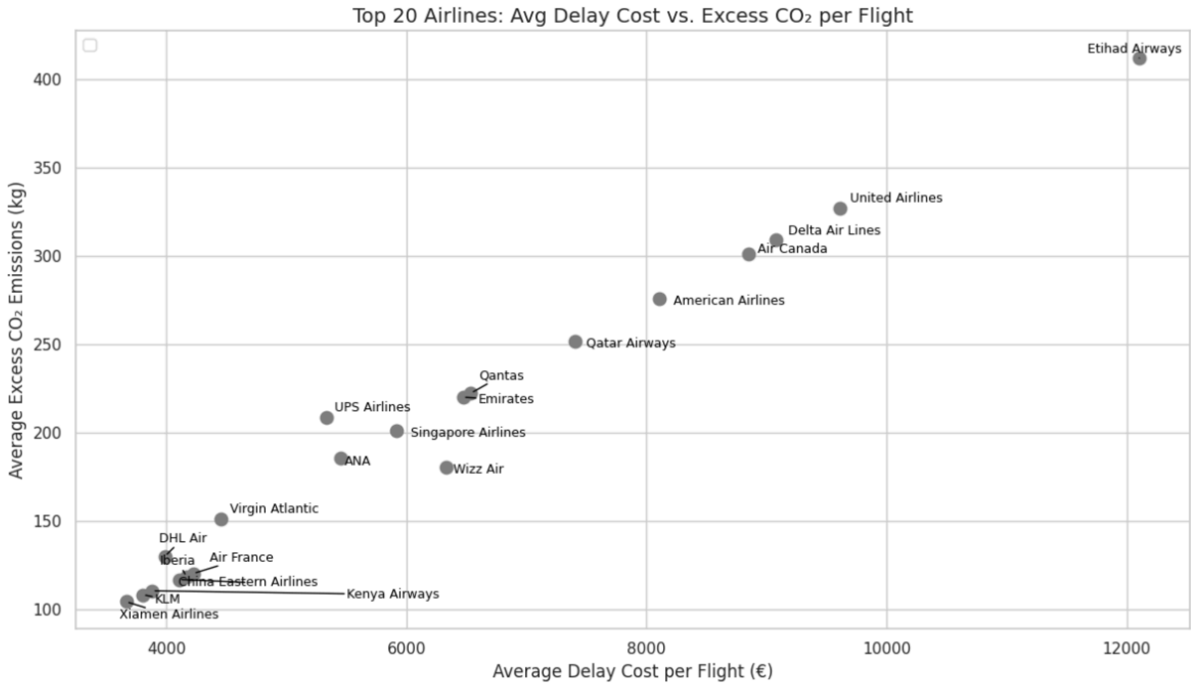


Figure 48 - Average delay cost vs. excess CO_2 emissions per flight for the top 20 airlines.

At Dublin Airport specifically, the monthly analysis in FIG 49 reveals that July and August experienced the highest total delay costs, aligning with peak summer travel demand. This pattern was also accounted for in the predictive modelling. Notably, this increase in cost did not always correspond to a higher number of delayed flights. In some months, relatively fewer delays led to high cumulative costs, likely due to more severe disruptions or the involvement of larger aircraft.

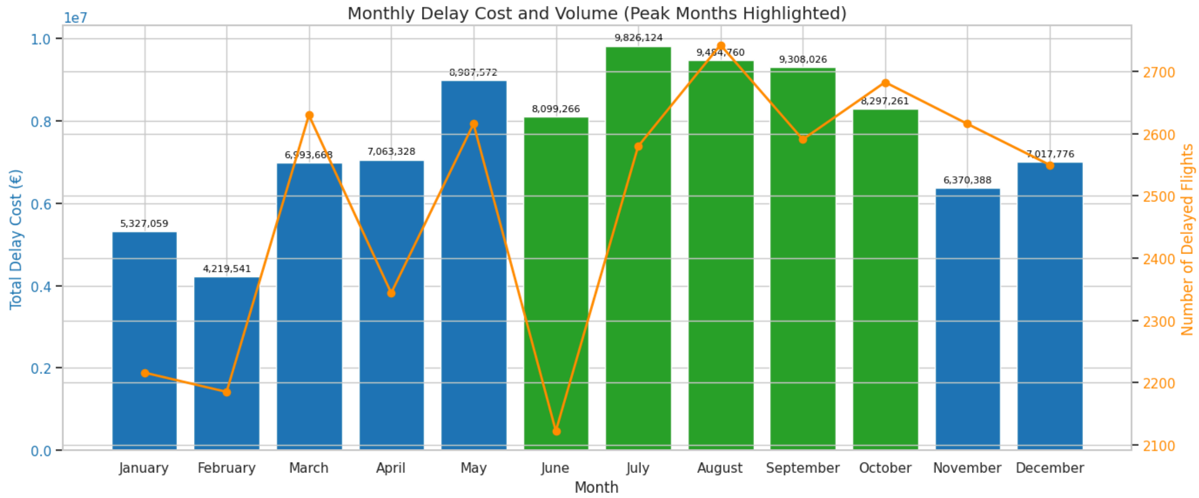


Figure 49 - Monthly delay cost and volume, with peak travel months highlighted in green.

Overall, the integration of financial metrics in this analysis provides a more transparent understanding of delay impacts at Dublin Airport. By accounting for aircraft size and seasonal dynamics, the findings move beyond basic statistics to uncover deeper operational inefficiencies. These insights support more informed decision-making not only for airport operators but also for airlines. As Dublin Airport continues to expand as a major international hub, such evidence-based approaches will be essential for balancing economic performance with long-term sustainability goals.

6.2 - Future work

This study opens several avenues for future exploration in this fascinating domain. Future research would benefit from integrating more precise aircraft-level data with regard to aircraft type and on the environmental side of the specific fuel burn rates and emissions. This would reduce the generalisation, even though it does highlight flaws, aircraft-specific metrics would enable a more accurate environmental assessment of delays.

One avenue not explored is the possibility of developing a real-time application of the optimally trained models, as it would provide strong practical value. Real-time predictions could aid Dublin Airport in operational decision making, resource allocation and passenger notifications to improve the overall capabilities of Dublin Airport.

A very positive next step would involve testing the broader applicability of the developed models, by applying this framework to other airports with similar operational characteristics to Dublin Airport. This would aid us when assessing whether the approach taken is case specific or generalisable across different locations. An obvious benefit of such comparative analysis would be to indicate both shared and unique delay drivers between airports. These developments would not only validate model performance across diverse settings but would also assist the development of more adaptable and scalable delay prediction systems.

A final adoption I would make for future work would be to integrate more temporal features that may not have been accounted for in the current model. These features would highlight industrial action, worker strikes and major events that can increase passenger volume and as a result, prolong delays.

6.3 – Review of Research Objectives

To investigate the influence of weather-related features on flight delays at Dublin Airport and assess how machine learning models capture and model these effects.

This study showed that after initial exploratory data analysis, weather variables (wind speed, visibility, temperature) had only a weak direct correlation with delay duration. This suggests that individually, these features may not influence delays directly. On the other hand, when incorporated into machine learning models, the same variables, especially wind speed and humidity combined, contribute meaningfully to the classification performance. This highlights how there may not be a simple correlation, but they interact in a more complex manner that models like XGBoost were able to capture and leverage for predictions. Therefore, weather conditions did play a role, maybe not as individual predictions but as a complex system influencing delays.

To evaluate the predictive performance of various machine learning algorithms in forecasting flight delays within a specific airport environment.

From the machine learning algorithms tested, stacking ensemble models combining Random Forest, XGBoost, and MLP (Neural Networks) proved the most effective in predicting significant delays. Final optimised models achieved 86% accuracy and an AUC score of 0.91, outperforming all individual models across various metrics. This supports previous literature in the domain of flight delay predictions. As individual models, Random Forest and XGBoost performed impressively, displaying tree-based ensemble methods and the ability to capture complex interactions between weather, temporal and operational features. Simpler models like logistic regression and Naive Bayes were found to have underperformed, defending previous literature in the need for more robust approaches in operational aviation research.

To quantify the contribution of flight delays at Dublin Airport to excess CO2 emissions using aircraft size-related features and operational data.

Delays have a direct impact on excess CO_2 emissions, as fuel burn time increases during extended taxiing and idling. This was quantified by categorising aircraft by size and applying EUROCONTROL fuel burn rates to estimate emissions per minute of delay. The analysis conducted showed a strong correlation, indicating that emissions rose proportionally with delayed duration. While, larger aircraft and peak-season delays contributed the most, highlighting the environmental cost of operational inefficiencies.

To identify and analyse the most influential features contributing to the prediction of flight delays.

The most influential features identified in this study were the hour of day, flight duration, and peak season indicators, all of which reflect operational strain and congestion. Flights departing during rush hours or in the summer months had a significantly higher probability of being delayed. Among weather variables, wind speed and temperature had the greatest impact, especially when combined with high humidity. SHAP analysis confirmed the dominance of temporal and operational features, while weather conditions contribute meaningfully through complex, non-linear interactions.

To explore how predictive modelling can inform and support delay mitigation strategies in airport operations.

The predictive models developed in this study lead to many practical delay mitigation strategies by identifying key risk factors in delayed flights, such as temporal factors, flight durations and some adverse weather scenarios. These insights found at Dublin Airport allow airport operators to implement smarter scheduling and allocate resources more effectively during peak periods. This study enables proactive interventions before disruptions escalate in predictable situations. Furthermore, by quantifying financial and environmental cost delays for Dublin Airport stakeholders, the models can support sustainability driven decisions, aiding the reduction of fuel waste and emissions through more efficient operations.

6.4 - Comparison with Existing Literature

To contextualise the results of this study, it is crucial to compare them with previous research on flight delay prediction using machine learning approaches. This is a well-established and actively researched domain, with studies employing a wide range of algorithms, data types, and modelling strategies. Performance outcomes often vary depending on the features selected and the geographical and operational characteristics of the airport in question. As such, drawing comparisons between Dublin Airport and similar international airports provides valuable insight. This section critically examines how the findings of the current study align with or diverge from prior work, focusing on model performance, dataset design, the role of weather variables, and the application of interpretability techniques.

The results of this study are consistent with previous literature in highlighting the superior performance of ensemble-based models for flight delay classification. Through multiple studies, methods such as Random Forest, XGBoost, and Stacking have demonstrated consistent advantages over traditional classifiers such as Logistic Regression and Naive Bayes, reflecting the findings of this research. Yi et al. (2021) conducted a detailed analysis using a Stacking model composed of Random Forest, Logistic Regression, Naive Bayes, and K-Nearest Neighbours (KNN), with Logistic Regression serving as the meta-learner. Applied to data from Boston Logan International Airport, their model achieved an AUC score of 0.823 for departure delays, with Random Forest also performing strongly as an individual classifier. These results closely mirror those of the present study, in which a similarly structured Stacking model achieved an AUC of 0.91, reinforcing the robustness of ensemble methods when applied to weather-affected datasets. Similarly, a study by Mtimkulu and Maphosa (2023) further confirmed the continuous advantages of ensemble techniques over traditional classifiers in flight delay prediction tasks, aligning closely with the findings of this research.

Similarly, Hatıpoğlu and Tosun (2024) evaluated a wide range of machine learning algorithms on flight data from a Turkish airport, including XGBoost, LightGBM, CatBoost, and Artificial Neural Networks. They found that XGBoost consistently delivered the highest performance, with an accuracy of 80%, particularly when SMOTE was used to address class imbalance and hyperparameter tuning was applied. The present study echoes these findings by also demonstrating the effectiveness of boosting methods in handling imbalanced classification tasks, a common challenge in flight delay datasets. Furthermore, the implementation of hyperparameter tuning contributed significantly to model optimisation, further aligning this research with best practices. One contrast to their findings, where weather variables had minimal predictive impact, we can see from the XGBoost SHAP values that wind speed, temperature and humidity have an impact on flight delay classification. This is due to the complex relationships within the dataset and due to the geolocation of Dublin in comparison to Turkey.

A comparison approach to this study is the deep learning approach explored by Schultz et al (2021), who applied CNNs and RNNs to classify weather impacts on airport performance across major European hubs. One similarity between our research was the use of METAR data reports, providing accurate weather updates. Schultz achieved a classification accuracy of 90% at Gatwick Airport, outperforming current models. This study found that simpler neural networks (MLP, CNN) underperformed compared to the tree-based models. This difference could be attributed to the differences in data used as Schultz's model leveraged weather severity levels categorised by the ATMAP algorithm whereas this work focused on static flight level data. One limitation to this approach is the limited interpretability of deep architectures, Reitmann et al (2019) seconded this limitation in their classification of airport performance. In contrast, this study addresses the limitation noted by integrating SHAP values, significantly enhancing the interpretability and transparency of model predictions.

Another key point of contrast is the data scope of other research in this domain. Previous studies conducted by Hatipoglu and Tosun (2024), Yi et al (2021), and Schultz et al (2021) rely on large, national or multi airport datasets. Kim and Park (2024) are a prime example of this as they applied various machine learning and deep learning methods across three major international airports: Incheon (ICN), JFK and Midway (MDW). They achieved an accuracy of 85.2% using their LSTM (Type of RNN) model, indicating strong predictive performance, but they did not explore model integration, such as stacking. Another limitation to this research was that it did not offer airport specific modelling, which is in direct contrast, the scope of this research is exclusively on Dublin Airport, aiming to conduct context specific analysis tailored to local weather and operational patterns. Schultz et al (2019) stated "the model must be adaptive" highlighting the importance of regional performance classification, therefore supporting the value of this localised model developed in this study of Dublin Airport.

Finally, the report by Donoghue and Cole (2018) provides a useful benchmark for traditional machine learning approaches in the context of flight delay prediction. Their study utilised a logistic regression model on historical U.S. network delay data and reported an AUC score of 0.69, highlighting the limitations of traditional linear classifiers in capturing the complex, nonlinear relationships. In contrast, the present study demonstrates a significant performance improvement through the application of a stacking ensemble, which achieved an AUC of 0.91. This result emphasises the added value of ensemble learning techniques and the incorporation of richer feature sets, particularly weather-related variables, in enhancing model accuracy for airport-specific delay prediction (Donoghue & Cole, 2018)

In conclusion, the findings of this study are broadly consistent with current literature, proving the strength of ensemble-based learning models for flight delay predictions. However, this study fills the research gap as it contributes uniquely through its exclusive focus on Dublin Airport, the integration of localised weather variables and increased explainability techniques through the application of SHAP. As mentioned earlier, many studies rely on large, multi-airport datasets or deep learning models, leaving limited interpretability. The present study highlights how well-structured ensemble modelling, when applied in a context specific setting, can achieve an impressive performance while maintaining interpretability and real-world applications.

6.5 - Limitations

A key limitation in this research was the availability and balance of aviation datasets. Comprehensive flight data is difficult to obtain, particularly when constrained by financial and licensing barriers, which was a common theme in the beginning. As a result, this study relied on a random sample of 100 flights per day over a one-year period. While this approach provides a manageable and representative subset, it inherently carries the risk of class imbalance, especially for rare events such as extreme delays. Random sampling, although practical, may inadvertently favour certain outcomes or patterns, reducing the diversity of the dataset and introducing bias.

To mitigate this, SMOTE (which was mentioned earlier) was applied during the training phase to address class imbalance. However, while SMOTE helps in creating a more balanced training set, it does not fully eliminate the limitations imposed by an incomplete or biased dataset.

Another limitation was the reliance on raw weather features rather than utilising derived metrics or clustered weather conditions, as seen in other studies. The ATMAP (Attribute Mapping) approach, which facilitates feature engineering and transformation, was not supported in the available infrastructure. This may have limited the model's ability to capture more abstract or non-linear interactions between weather variables that could influence flight delays.

Another limitation in this study was the lack of detailed aircraft type information in the original dataset. To address this, aircraft types were manually mapped based on the most common models flown by each airline out of Dublin Airport. While this provided a reasonable approximation and allowed for the inclusion of aircraft-related factors, it does introduce a degree of generalisation.

Lastly, external factors influencing flight delays, such as airline scheduling practices, air traffic control decisions, or airport-specific operational events were not included in the dataset. Their exclusion may have reduced the explanatory power of the models and limited the scope of insights derived from the analysis.

Chapter 7 – Conclusion

7.1 Conclusion

This study confirms the research question that machine learning can be used to accurately predict flight delays at Dublin Airport, while providing actionable insights into the financial and environmental consequences of delays. By combining advanced classification models with explainability tools and impact metrics, the research goes beyond simple prediction to explore the broader financial and environmental implications of disruption.

The machine learning results demonstrate that ensemble models, particularly stacking, deliver the highest predictive accuracy and reliability at recognising minority classes. These models were successfully able to capture complex patterns shaped by operational, temporal and meteorological variables including weather conditions like temperature, wind speed and humidity. Unlike previous studies, SHAP analysis enabled a transparent interpretation of the predictions, allowing for actionable insights for airport and airline stakeholders.

Another valuable dimension of this study is the incorporation of size base financial cost modelling and CO2 emission estimates. The findings show that delays are not only frequent but also costly, economically and environmentally. A key similarity between discoveries was that large aircraft, long-haul carriers and peak season operations emerged as key drivers of both the excess costs and emissions. These findings have direct real-world applicability for Dublin Airport, similarly to global efforts where operational efficiency and sustainability are becoming increasingly interlinked goals.

Ultimately, this research delivers a scalable, interpretable, and airport-specific framework for flight delay prediction. It enables more efficient resource planning, promotes sustainable airport operations, and strengthens resilience to weather-related disruptions. As Dublin Airport continues its expansion as a key international hub, the insights developed in this study offer a timely and practical solution to the evolving challenges of aviation management.

This dissertation presents a cohesive journey from identifying a critical operational challenge at Dublin Airport to developing and validating a predictive framework that integrates weather data, machine learning, economic and environmental impact analysis. By linking exploratory findings with model outputs and interpreting results through SHAP, this research demonstrates the value of data-driven approaches to enhance decision-making and sustainability in aviation.

Bibliography

AirHelp (2023). *In numbers: The economic impact of flight disruptions*, 16 November. Available at: https://www.airhelp.com/en/blog/in-numbers-the-economic-impact-of-flight-disruptions/

Anupkumar, A. (2023). *Investigating the costs and economic impact of flight delays in the aviation industry and the potential strategies for reduction*. https://scholarworks.lib.csusb.edu/etd/1653

Baranishyn, M., Cudmore, B. A., & Fletcher, T. (2010). *Customer service in the face of flight delays*. *Journal of Vacation Marketing, 16*(3), 201–215. https://doi.org/10.1177/1356766710373681

Boggavarapu, R., Agarwal, P., & Rohith Kumar, D. H. (2019). *Aviation delay estimation using deep learning*. In *Proceedings of the 2019 4th International Conference on Information Systems and Computer Networks (ISCON 2019)* (pp. 689–693). IEEE. https://www.researchgate.net/publication/339981003_Aviation_Delay_Estimation_using_Deep_Learning

Breiman, L. (2001). *Random forests*. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Cole, S., & Donoghue, T. (2016). *Predicting departure delays of US domestic flights: CSE 258 Assignment 2*. University of California, San Diego. https://srcole.github.io/assets/flight_delay/report.pdf

daa. (2025). *Annual Report & Accounts 2024*. Dublin Airport Authority. https://www.daa.ie

Dimitriou, D., & Karagkouni, A. (2022). *Airports' sustainability strategy: Evaluation framework upon environmental awareness*. *Frontiers in Sustainability*, *3*, Article 880718. https://doi.org/10.3389/frsus.2022.880718

Efthymiou, M., Njoya, E. T., Lo, P. L., Papatheodorou, A., & Randall, D. (2019). *The impact of delays on customers' satisfaction: An empirical analysis of British Airways on-time performance at Heathrow Airport*. *Journal of Aerospace Technology and Management, 11*, e201911. https://jatm.com.br/jatm/article/view/977/737

ENVISA. (2006). *Project GAES – Environmental impact of delay* (EEC/SEE/2006/006). EUROCONTROL Experimental Centre. https://www.eurocontrol.int/sites/default/files/library/036_Environmental_Impact_of_Delay.pdf

EUROCONTROL. (2024, January 23). *EUROCONTROL European aviation overview – 2024 in review*. https://www.eurocontrol.int/sites/default/files/2025-01/eurocontrol-european-aviation-overview-20250123-2024-review.pdf

Hatıpoğlu, I., & Tosun, Ö. (2024). *Predictive modeling of flight delays at an airport using machine learning methods*. *Applied Sciences*, *14*(13), 5472. https://doi.org/10.3390/app14135472

Hossin, M., & Sulaiman, M. N. (2015). *A review on evaluation metrics for data classification evaluations*. *International Journal of Data Mining & Knowledge Management Process*, *5*(2), 1–11. Retrieved March 23, 2025, from https://doi.org/10.5121/ijdkp.2015.5201

ICAO. (2022). *2022 Environmental Report: Innovation for a Green Transition*. International Civil Aviation Organisation. https://www.icao.int/environmental-protection/Pages/env-report.aspx

IEEE. (2017). *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*. IEEE. https://doi.org/10.1109/DASC.2017.8102016

Jiang, W., Chen, Z., Xiang, Y., Shao, D., Ma, L., & Zhang, J. (2019). *SSEM: A novel self-adaptive stacking ensemble model for classification*. https://www.researchgate.net/publication/335156833_SSEM_A_Novel_Self-Adaptive_Stacking_Ensemble_Model_for_Classification

Kennon, P., Hazel, R. A., Ford, E. K., & Hargrove, B. (2013). *Preparing peak period and operational profiles: Guidebook* (ACRP Report 82). Transportation Research Board. https://doi.org/10.17226/22646

Khaksar, H., & Sheikholeslami, A. (2019). *Airline delay prediction by machine learning algorithms*. *Scientia Iranica*, *26*(5 A), 2689–2702. https://doi.org/10.24200/sci.2017.20020

Khan, M., Hooda, B. K., Gaur, A., Singh, V., Jindal, Y., Tanwar, H., Sharma, S., Sheoran, S., Vishwakarma, D. K., Khalid, M., Albakri, G. S., Alreshidi, M. A., Choi, J. R., & Yadav, K. K. (2024). *Ensemble and optimisation algorithm in support vector machines for classification of wheat genotypes*. *Scientific Reports*, *14*(1), Article 22728. https://doi.org/10.1038/s41598-024-72056-0

Khan, R., Akbar, S., & Zahed, T. A. (2022). *Flight delay prediction based on gradient boosting ensemble techniques*. In *Proceedings of the 2022 16th International Conference on Open Source Systems and Technologies (ICOSST 2022)*. IEEE. https://doi.org/10.1109/ICOSST57195.2022.10016828

Kim, S., & Park, E. (2024). *Prediction of flight departure delays caused by weather conditions adopting data-driven approaches*. *Journal of Big Data, 11*(1), Article 1. https://doi.org/10.1186/s40537-023-00867-5

Kulesa, G. (2003). *Weather and aviation: How does weather affect the safety and operations of airports and aviation, and how does FAA work to manage weather-related effects?* U.S. Department of Transportation Center for Climate Change and Environmental Forecasting.

Li, X., Chen, X., & Liu, Z. (2022). *Research on construction and development of green airport*. In *Proceedings of the 2022 International Conference on Urban Planning and Regional Economy (UPRE 2022)* (pp. 434–437). Atlantis Press. https://doi.org/10.2991/aebmr.k.220502.077

Lui, M. (2014). *Complete Decoding and Reporting of Aviation Routine Weather Reports (METARs)* https://ntrs.nasa.gov/citations/20150000833

Midtfjord, A. D., De Bin, R., & Huseby, A. B. (2022). *A decision support system for safer airplane landings: Predicting runway conditions using XGBoost and explainable AI. Cold Regions Science and Technology*, *199*, 103556. https://doi.org/10.1016/j.coldregions.2022.103556

Mijwil, M. M. (2021). Artificial Neural Networks Advantages and Disadvantages. *Mesopotamian Journal of Big Data*, *2021*, 29–31. https://doi.org/10.58496/mjbd/2021/006

Mtimkulu, Z., & Maphosa, M. (2023). *Flight delay prediction using machine learning: A comparative study of ensemble techniques*. In *Proceedings of the 2023 International Conference on Artificial Intelligence and Its Applications* (pp. 212–218). https://doi.org/10.59200/icarti.2023.030

Muraina, I. O., Adesanya, O., & Abam, S. O. (n.d.). *Data analytics evaluation metrics essentials: Measuring model performance in classification and regression*. Retrieved March 23, 2025, from https://www.researchgate.net/publication/372883214_DATA_ANALYTICS_EVALUATION_METRICS_ESSENTIALS_MEASURING_MODEL_PERFORMANCE_IN_CLASSIFICATION_AND_REGRESSION

Naskath, J., Sivakamasundari, G., & Begum, A. A. S. (2023). *A study on different deep learning algorithms used in deep neural nets: MLP, SOM and DBN. Wireless Personal Communications*, *128*(4), 2913–2936. https://doi.org/10.1007/s11277-022-10079-4

Parrilla Gutiérrez, J. M. (2010). *Support vector machines: Similarity functions to work with heterogeneous data and classifying documents* (Master's thesis). Universitat Politècnica de Catalunya. https://upcommons.upc.edu/handle/2099.1/11809

Performance Review Commission. (2012). *Performance review report: An assessment of air traffic management in Europe during the calendar year 2011*. EUROCONTROL. https://www.eurocontrol.int/sites/default/files/publication/files/prr-2011.pdf

Popescu, M.-C., & Balas, V. E. (2009). *Multilayer perceptron and neural networks*. *WSEAS Transactions on Circuits and Systems*, *8*(7), 579–588 https://www.researchgate.net/publication/228340819_Multilayer_perceptron_and_neural_networks

Purushothaman, S., Sankar, S., Parthiban, S., Kumaran, G., Pradeepkumar, S., & Manikavelan, D. (2024). Flight Delay Prediction for Air Travel Management Using Artificial Neural Network Compared with Auto Regressive Integrated Moving Average Algorithm. https://www.researchgate.net/publication/380829442_Flight_Delay_Prediction_for_Air_Travel_Management_Using_Artificial_Neural_Network_Compared_with_Auto_Regressive_Integrated_Moving_Average_Algorithm

Reitmann, S., Alam, S., & Schultz, M. (2019). *Advanced quantification of weather impact on air traffic management: ATMAP 2.0 – Intelligent weather categorisation with machine learning*. https://www.researchgate.net/publication/331651855_Advanced_Quantification_of_Weather_Impact_on_Air_Traffic_Management_-_Intelligent_Weather_Categorization_with_Machine_Learning

Schultz, M., Lorenz, S., Schmitz, R., & Delgado, L. (2018). *Weather impact on airport performance*. *Aerospace*, *5*(4), Article 109. https://doi.org/10.3390/aerospace5040109

Schultz, M., Reitmann, S., & Alam, S. (2021). *Predictive classification and understanding of weather impact on airport performance through machine learning*. *Transportation Research Part C: Emerging Technologies*, *131*, 103119. https://doi.org/10.1016/j.trc.2021.103119

Schwarz, P. (2022). Stacking ensemble classification applied to US flight delay prediction during covid-19 pandemic.

Shehab, M., Moshammer, K., Franke, M., & Zondervan, E. (2023). *Analysis of the potential of meeting the EU's sustainable aviation fuel targets in 2030 and 2050*. *Sustainability*, *15*(12), 9266. https://doi.org/10.3390/su15129266

Song, M., Wang, J., & Li, R. (2024). *The importance of weather factors in the resilience of airport flight operations based on Kolmogorov–Arnold Networks (KANs)*. *Applied Sciences*, *14*(19), 8938. https://doi.org/10.3390/app14198938

Soofi, A. A., & Awan, A. (2017). *Classification techniques in machine learning: Applications and issues*. *Journal of Basic & Applied Sciences*, *13*, 459–465. https://doi.org/10.6000/1927-5129.2017.13.76

Sternberg, A., Soares, J. A., Carvalho, D., & Ogasawara, E. (2021). *A review on flight delay prediction*. *Transport Reviews*, *41*(4), 499–528. https://doi.org/10.1080/01441647.2020.1861123

Sujay, V., Lalitha, S., Sreedhar, S. S., Omotunde, H., Vijaya, S., & Mery, K. P. (2024). *Enhancing hyperparameters for improved flight delay prediction using machine learning algorithms*. In *Proceedings of the 2nd International Conference on Integrated Circuits and Communication Systems (ICICACS 2024)*. IEEE. https://www.researchgate.net/publication/379935958_Enhancing_Hyperparameters_for_Improved_Flight_Delay_Prediction_Using_Machine_Learning_Algorithms

Tang, Y. (2021). *Airline flight delay prediction using machine learning models*. In *Proceedings of the 2021 International Conference on Artificial Intelligence and Computer Science (AICS 2021)* (pp. 151–154). ACM. https://doi.org/10.1145/3497701.3497725

Venkatesh, V., Arya, A., Agarwal, P., Lakshmi, S., & Balana, S. (2017). Iterative Machine and Deep Learning Approach for Aviation Delay Prediction. 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON), 562–567. https://www.researchgate.net/publication/322412104_Iterative_machine_and_deep_learning_approach_for_aviation_delay_prediction

Wang, X., Wang, Z., Wan, L., & Tian, Y. (2022). *Prediction of flight delays at Beijing Capital International Airport based on ensemble methods*. *Applied Sciences*, *12*(20), Article 10621. https://doi.org/10.3390/app122010621

Wang, Y. (n.d.). *Prediction of Weather Impacted Airport Capacity using Ensemble Learning*. https://ntrs.nasa.gov/api/citations/20140008304/downloads/20140008304.pdf

Yaşar Dinçer, F. C., Yirmibeşoğlu, G., Bilişli, Y., Arık, E., & Akgün, H. (2024). *Trends and emerging research directions of sustainable aviation: A bibliometric analysis*. *Heliyon*, *10*(11), e32306. https://doi.org/10.1016/j.heliyon.2024.e32306

Ye, B., Liu, B., Tian, Y., & Wan, L. (2020). *A methodology for predicting aggregate flight departure delays in airports based on supervised learning*. *Sustainability*, *12*(7), 2749. https://doi.org/10.3390/su12072749

Yi, J., Zhang, H., Liu, H., Zhong, G., & Li, G. (2021). *Flight delay classification prediction based on stacking algorithm*. *Journal of Advanced Transportation*, *2021*, Article 4292778. https://doi.org/10.1155/2021/4292778

Appendix A - Modelling Dataset Features

Variable	Description
Flight Number	Unique Identifier
Airline	Carrier operating flight
Aircraft Model	Specific Aircraft type
Departure Terminal	Terminal flight departs
Departure Gate	Gate flight departs
Arrival Airport	Destination
Aircraft Size	Grouped size category
Flight Duration	Duration in hours
Peak hour	Binary variable if flights depart during peak hours
Peak days	Binary variable if flights depart during peak days
Peak season	Binary variable if flights depart during peak season
Wind Speed (KT)	Surface wind speed at time of departure
Visibility (m)	Visibility at time of departure
Temperature (C)	Temperature at time of departure
Dew Point (C)	Dew point at time of departure
Pressure (hPa)	Pressure at time of departure
Humidity (%)	Humidity at time of departure
Hour of day	Departure Hour
Day of week	Day of the week
Delayed (Target)	Delayed vs On time

Appendix B - Link to GitHub Repository

This is the link to all the code used in this research:
https://github.com/abolger13/Thesis-Code.git