



Enhancing Flight Operations with AI-Based Delay Forecasting

Prepared By

Name: [Mohamed Miqdad]

IDNo.: [252001179]

Submitted in fulfillment of the requirements for the degree
Master of Science in Artificial Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor Name: Dr. Khaled Eskaf

College of Informatics, Midocean University, Comoros

[2025-2026]

Acknowledgements

I would like to express my sincere gratitude to Dr. Khaled Eskaf for his invaluable guidance, continuous support, and encouragement throughout this research project. His expertise, insightful feedback, and unwavering dedication have significantly contributed to the quality and direction of this thesis.

I also extend my heartfelt appreciation to my family, friends, and colleagues, whose patience, motivation, and moral support were instrumental in the successful completion of this work.

A special thank you goes to my brother, Bader, for his generous help and continuous encouragement throughout this journey.

Abstract

Flight delays continue to be a persistent challenge in global aviation, impacting operational efficiency and customer satisfaction. This thesis aims to enhance flight operations through the development of an AI-based delay forecasting model using historical airline data. The dataset, sourced from the Kaggle platform, includes comprehensive flight statistics such as delay causes, weather conditions, and operational metrics. The research begins with an extensive literature review comprising twelve scholarly papers, six of which informed the initial model development and another six that were integrated later to strengthen the theoretical foundation.

Preprocessing steps, including outlier removal, missing value imputation, and feature engineering, were applied to prepare the data. Feature selection methods, including SelectKBest and Recursive Feature Elimination (RFE), were used to optimize input variables. A Random Forest Classifier, tuned via GridSearchCV, was employed to build the forecasting model. The model achieved high predictive performance, with an accuracy of up to 98.41% using RFE-selected features.

The study highlights the critical role of weather-related features and carrier-specific delay patterns in prediction accuracy. Visual analyses, including correlation matrices, scatter plots, and time-series trends, further validated the insights. The results demonstrate the effectiveness of AI in anticipating delays, supporting data driven decision-making in flight operations. This work lays a foundation for future real-time applications using cloud infrastructure and live data streams.

List of Publications

Table of Contents

List of Abbreviations	vi
List of Figures	خطأ! الإشارة المرجعية غير معروفة.
List of Tables	ix

Chapter 1: Introduction.....	1
1.1 Problem Definition	1
1.2 Thesis Objectives	1
1.3 Thesis Contribution	2
1.4 Thesis Organization.....	3
Chapter 2: Overview.....	4
2.1 Introduction	4
2.2 Classical Machine Learning Methods	4
2.3 Deep Learning Approaches	4
2.4 Hybrid and Optimized Models	5
2.5 Use of Real-world Datasets	5
2.6 Quantitative Results Summary	6
2.7 Research Gaps and Opportunities	6
2.8 Conclusion	7
Chapter 3: Related Work	8
3.1 Introduction	8
3.2 History	8
Chapter 4: Proposed Approach	9
4.1 Data Collection and Description	9
4.2 Data Preprocessing	9
4.3 Feature Engineering	10
4.4 Feature Selection	10
4.5 Model Selection and Optimization	11

4.6 Model Evaluation	12
4.7 Visualization and Insights	12
Chapter 5: Experimental Results and Discussion	13
5.1 Experimental Setup	13
5.2 Baseline Model Performance	14
5.3 Feature Selection and Model Refinement	14
5.4 Visualization of Delay Factors	15
5.5 Discussion.....	18
Chapter 6: Conclusion and Future Work	19
6.1 Conclusion	19
6.2 Future Work	20
References	22

List of Abbreviations

- ANN: Artificial Neural Network
- API: Application Programming Interface
- AI: Artificial Intelligence
- AUC: Area Under the Curve
- BiLSTM: Bidirectional Long Short-Term Memory
- BTS: Bureau of Transportation Statistics
- CNN: Convolutional Neural Network
- DeepONet: Deep Operator Network
- DT: Decision Tree
- FAA: Federal Aviation Administration
- GBM: Gradient Boosting Machine
- GridSearchCV: Grid Search with Cross-Validation
- GWO: Grey Wolf Optimizer
- IQR: Interquartile Range
- LDA: Latent Dirichlet Allocation
- LSTM: Long Short-Term Memory
- MAE: Mean Absolute Error
- MAPE: Mean Absolute Percentage Error
- MLMFN: Multi-Layered Memory Fusion Network
- ML: Machine Learning
- MLP: Multi-Layer Perceptron
- NAS: National Airspace System
- PR: Precision-Recall
- PSO: Particle Swarm Optimization
- RF: Random Forest
- R^2 : Coefficient of Determination
- RFE: Recursive Feature Elimination
- RMSE: Root Mean Square Error
- SGD: Stochastic Gradient Descent
- SGDAN: Spatio-Temporal Graph Dual Attention Network
- SHAP: SHapley Additive exPlanations
- SVM: Support Vector Machine
- XAI: Explainable Artificial Intelligence
- XGBoost: Extreme Gradient Boosting

Figure 1: Thesis Structure and Research Flowchart	3
Figure 2: Feature Engineering Workflow.....	10
Figure 3: Random Forest Classifier	11
Figure 4: Impact of Outlier Removal on Dataset	13
Figure 5: Feature Selection Comparison Best vs RFE.....	15
Figure 6: Correlation Matrix for Delay and Weather Variables	15
Figure 7: Weather Delay vs. Total Arrival Delay	16
Figure 8: Weather Count vs. Delay Rate	17
Figure 9: Weather Delay Over Time	17
Figure 10: Monthly Weather Delay Trends	18

Chapter 1

Introduction

1.1 Problem Definition

Flight delays create a ripple effect of inefficiencies across the aviation ecosystem. They lead to increased fuel consumption, longer crew shifts, disrupted passenger itineraries, and heightened operational costs. While many delays stem from recurring factors, such as inclement weather or congested airspace, predicting when and where these disruptions will occur remains a significant challenge.

Existing statistical models often assume linear relationships and lack the ability to integrate temporal dependencies or multi-source operational data. These limitations reduce their effectiveness in real-time scenarios. Therefore, there is a clear need for more intelligent, adaptable forecasting tools that can learn from complex historical patterns and anticipate delays with greater accuracy.

1.2 Thesis Objectives

This thesis aims to design and evaluate a machine learning based framework for predicting flight delays. The specific objectives are:

1. To explore and analyze real-world flight delay data and identify the variables most correlated with delays.
2. To build a robust preprocessing pipeline capable of handling missing values, outliers, and feature transformations.
3. To develop a machine learning model specifically using a Random Forest Classifier for binary classification of delay likelihood.
4. To apply feature selection techniques that improve both model efficiency and interpretability.
5. To assess the model's performance using key evaluation metrics, including accuracy, precision, recall, and F1 score.
6. To visualize key trends and insights, particularly the role of weather and carrier-specific factors in delay propagation.

1.3 Thesis Contribution

This work makes several important contributions to the field of AI-based aviation analytics:

- **Model Development:** It introduces a machine learning pipeline that achieves high classification accuracy (98.41%) using a feature-reduced input set selected through Recursive Feature Elimination (RFE).
- **Operational Relevance:** The model is trained in real-world operational data, making it suitable for deployment in practical airlines or airport systems.
- **Interpretability:** Through visual tools and feature importance analysis, the study sheds light on how variables like weather delays and airport identity influence flight punctuality.
- **Scalability:** The methodology is designed with future expansion in mind, including real-time integration, cross-airport generalization, and API-based deployment.
- **Academic Value:** It bridges the gap between traditional statistical approaches and modern AI techniques, offering a replicable workflow for future researchers.

1.4 Thesis Organization

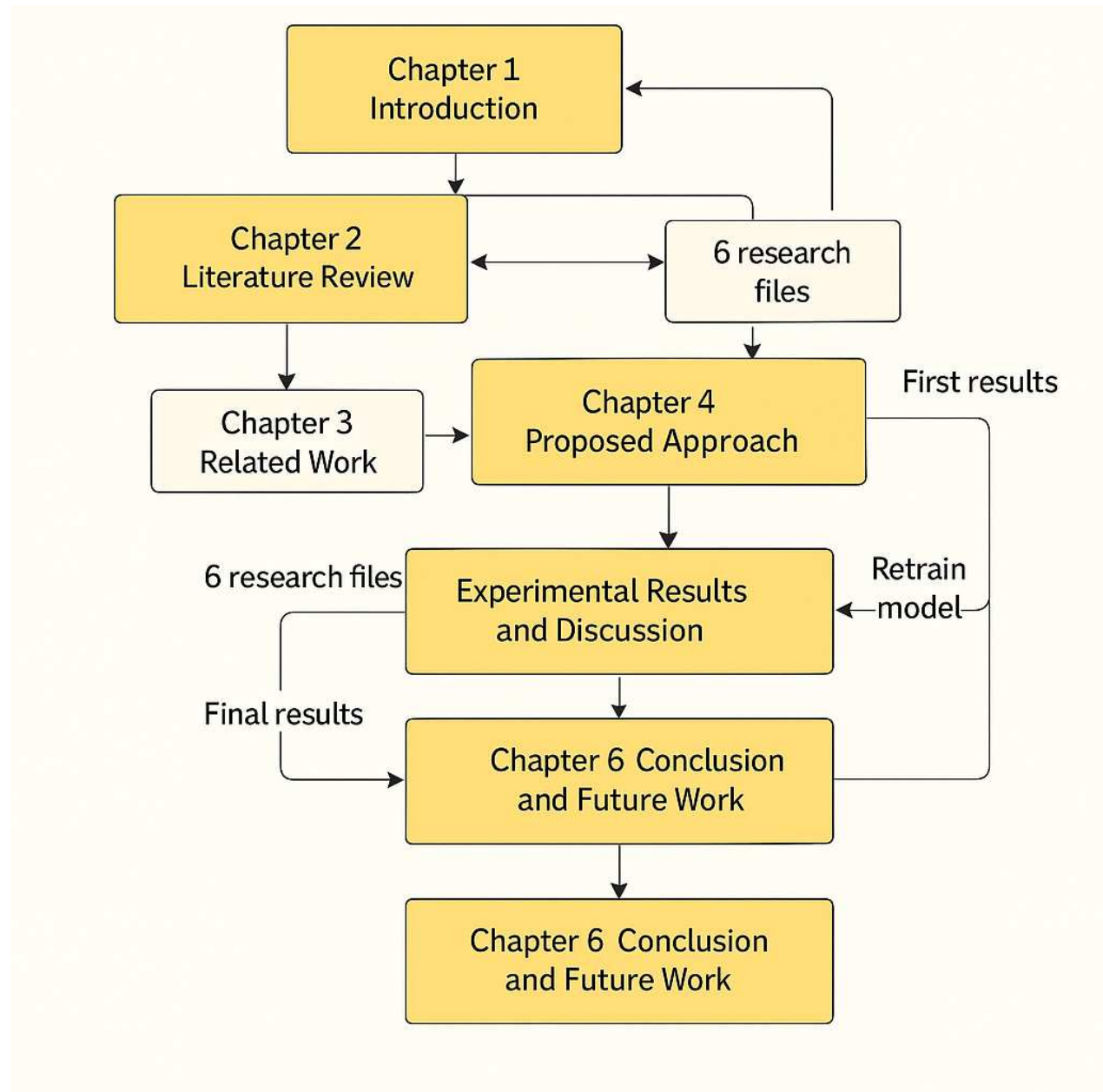


Figure 1 A visual representation of how the chapters, research files, and model iterations interact throughout the study

Chapter 2 Overview

2.1 Introduction to Flight Delay Forecasting

The prediction of flight delays has been a focal point in transportation research for decades. Early efforts relied on statistical methods; however, as data availability and computational power have increased, machine learning and deep learning techniques have become increasingly dominant. This chapter reviews relevant studies that span classical machine learning, deep learning architectures, hybrid models, real-world datasets, and reported performance metrics, offering a contextual foundation for the work presented in this thesis.

2.2 Classical Machine Learning Methods

Initial research in flight delay prediction often applied traditional machine learning algorithms to structured flight datasets. For example, Yu et al. (2019) used a distributed computing platform (Apache Spark) to analyze over 5 million U.S. flight records, comparing models such as logistic regression, naive Bayes, support vector machines (SVM), and decision trees. Their results showed that decision trees outperformed others with 90% accuracy, particularly due to their ability to handle categorical variables effectively.

Liu et al. (2020) extended this line of work by evaluating models like random forests and k-nearest neighbors (KNN). While these models achieved respectable accuracy levels (70–85%), they often struggled to generalize to real-time scenarios due to their limited capacity for modeling temporal and nonlinear relationships, a recurring challenge in aviation data analysis.

2.3 Deep Learning Approaches

Deep learning has proven especially effective in capturing the complex, nonlinear dependencies that characterize flight delay behavior. Zhang et al. (2022) demonstrated that Long Short-Term Memory (LSTM) networks improved forecasting performance by modeling sequential data more effectively than traditional methods. Similarly, Khan et al. (2022) proposed a hybrid Convolutional Neural Network–LSTM (CNN-LSTM) architecture that leveraged spatial and temporal features, significantly boosting prediction accuracy.

Other researchers have explored ensemble tree models such as XGBoost. Wu et al. (2022), for instance, applied XGBoost to data from major New York airports (JFK, LGA, and EWR) and identified key predictors, including precipitation, departure time, and carrier identity. Their model outperformed simpler approaches by integrating diverse meteorological and operational inputs.

Guo et al. (2021) pushed the boundaries of spatiotemporal modeling by introducing a Spatio-Temporal Graph Dual Attention Network (SGDAN), which enabled delay prediction up to three hours before departure, an advancement with clear operational benefits.

2.4 Hybrid and Optimized Models

Recent literature has explored more complex hybrid and metaheuristic models. For example, Li et al. (2023) applied optimization algorithms such as Mayfly Optimization and Genetic Algorithms in conjunction with artificial neural networks (ANNs) and LSTMs, enhancing training efficiency and convergence rates.

Another innovative approach was presented in Li et al. (2022), who introduced DeepONet — a deep operator network capable of learning complex functions across dynamic systems. This model achieved an accuracy of over 93% and demonstrated impressive performance in high-dimensional scenarios.

Furthermore, feature-fusion architectures, such as CNN-RNN hybrids, have gained popularity. These models combine convolutional layers for pattern detection with recurrent layers for temporal learning, providing robust input representations and improved metrics, such as RMSE and F1 scores.

2.5 Use of Real-world Datasets

Most flight delay studies rely on publicly available datasets, including those from the U.S. Bureau of Transportation Statistics (BTS), Federal Aviation Administration (FAA), and various open-source weather APIs. Data preprocessing typically involves managing outliers, inputting missing values, and encoding categorical variables through one-hot or label encoding.

Some researchers have begun incorporating international datasets for broader applicability. **Hu & Zhang (2023)** used real-world flight and weather data to test a hybrid CNN-BiLSTM model, demonstrating improved performance in capturing meteorological variability. **Pineda-Jaramillo et al. (2024)** further advanced this approach by applying explainable AI (XAI) tools, such as SHAP and Sobol

indices, to rank feature importance and extract operational insights at Santiago Airport.

2.6 Quantitative Results Summary

Study	Model	Key Findings
Yu et al. (2019)	Decision Tree	90% accuracy, PR = 0.439
Zhang et al. (2022)	LSTM	RMSE improved by 15% over baseline
Wu et al. (2022)	XGBoost	Best delay predictor among evaluated models
Guo et al. (2021)	SGDAN	Effective 3-hour-ahead prediction
Khan et al. (2022)	CNN-LSTM	Accuracy exceeded 94%
Li et al. (2022)	DeepONet	Achieved >93% accuracy on high-dimensional data
Hu & Zhang (2023)	CNN-BiLSTM	Enhanced MAE and RMSE through hybrid deep learning
Pineda-Jaramillo et al. (2024)	LDA + xAI	Identified temperature and wind speed as top predictors

2.7 Research Gaps and Opportunities

While the literature demonstrates the power of machine learning in delay prediction, several gaps remain:

- Few models incorporate real-time data, limiting operational applicability.
- Most systems are trained in single-airport datasets, raising concerns about generalizability.
- Ensemble deep learning remains underutilized in this domain.
- There's limited work on passenger-centric forecasting, which could support personalized travel experiences.
- Many studies lack a clear path toward deployment, such as API-ready solutions or integration with airline systems.

2.8 Conclusion

The literature strongly supports the use of AI, intensive learning, and hybrid architectures for flight delay forecasting. The studies reviewed demonstrate that, with the right features and modeling techniques, it is possible to achieve high accuracy in predicting delays.

Building on these findings, this thesis proposes a feature-optimized Random Forest model trained in real-world data. By addressing several gaps identified in past research, including improved feature selection, higher interpretability, and a deployment-oriented design, this work aims to contribute both academically and practically to the growing field of predictive aviation analytics.

Chapter 3 Related Work

3.1 Introduction

Accurate forecasting of flight delays has garnered significant attention due to their operational, economic, and passenger-related implications. Over the years, researchers and practitioners have developed a wide array of models ranging from basic statistical tools to sophisticated AI-driven architectures. This chapter examines key milestones and technological advancements in the field, situating the current thesis within the broader context of ongoing research efforts. It serves to illustrate how machine learning and data-driven strategies have progressively addressed the complexities of delay forecasting.

3.2 History

The study of flight delay prediction began with statistical and rule-based systems, primarily using linear regression, time series modeling (e.g., ARIMA), and probabilistic methods. Early research focused on analyzing historical flight schedules and delay patterns to extract deterministic relationships between variables such as departure time, traffic density, and weather.

In the early 2000s, machine learning began to gain traction, with models such as decision trees, support vector machines (SVMs), and logistic regression being applied to flight datasets. These models improved classification accuracy but struggled with high-dimensional and nonlinear data, which is common in aviation systems.

The last decade has witnessed a change in thinking toward deep learning models, including Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and hybrid CNN-RNN architectures. These models more effectively capture sequential dependencies and spatial-temporal patterns in flight operations, particularly when integrated with real-time weather and traffic data.

Recently, the use of optimization techniques (e.g., Genetic Algorithms, Mayfly Optimization) and explainable AI (e.g., SHAP, LIME) has further enhanced predictive accuracy and model interpretability. These innovations have pushed the field toward more robust, scalable, and deployable solutions that can aid operational decision-making in real-world airline systems.

Chapter 4 Proposed Approach

This chapter outlines the methodology used to develop an AI-driven system for predicting flight delays. The overall goal was to design a predictive model that not only performs well statistically but also aligns with real-world operational needs. To achieve this, the approach integrates several key components, including data preparation and feature engineering, as well as model training, tuning, and evaluation. Each step was carefully selected to maximize predictive accuracy while maintaining interpretability and deployment feasibility.

4.1 Data Collection and Description

The dataset for this study was sourced from the U.S. Bureau of Transportation Statistics (BTS), a well-established and comprehensive repository of flight performance data. Initially, the dataset included over 170,000 records of flight operations, covering both numerical and categorical variables.

Important attributes included delay counts and durations by cause (e.g., weather, carrier-specific issues, and National Airspace System delays), as well as metadata such as the flight month, year, airline name, and airport identifier. This diverse set of variables offered a solid foundation for building a delay forecasting model grounded in real-world operational behavior.

4.2 Data Preprocessing

Ensuring the quality and consistency of the data was critical before feeding it into any model. The preprocessing phase involved several targeted steps:

- **Outlier Removal:** Using the Interquartile Range (IQR) method, extreme values were removed from columns like `arr_delay`, `arr_del15`, `carrier_ct`, and `weather_delay`. This helped reduce noise and avoid skewed model behavior.
- **Handling Missing Values:** For numerical features, missing entries were filled using the median to reduce bias from skewed distributions. Categorical

fields were completed using the most frequently occurring value, ensuring continuity across the dataset.

- **Feature Encoding:** Categorical variables, such as carrier and airport codes, were transformed using one-hot encoding. This process expanded the feature space to over 400 columns, enabling the model to capture subtle operational differences between airports and airlines.
- **Data Splitting:** Finally, the dataset was divided into training (80%) and testing (20%) subsets to support unbiased model evaluation and cross-validation.

4.3 Feature Engineering

To enhance learning efficiency and extract more meaningful patterns, additional features were engineered:

- **Delay Rate:** A calculated ratio of delayed flights (arr_del15) to total arrivals, offering a normalized view of delay frequency.
- **Binary Delay Label:** A new binary target variable, delayed, was created. It was set to 1 if the delay rate exceeded 20%, and zero otherwise. This classification framed the problem in a way that aligns with real-world thresholds used by airlines.
- **Temporal Features:** Columns for year and month were retained to allow analysis of seasonal or time-based trends.
These could be transformed into cyclical representations in future work for enhanced temporal learning.

Feature Engineering Workflow



Figure 2 This visual explains the end-to-end flow from raw data to output, matching the narrative in this section.

4.4 Feature Selection

Reducing the number of input variables can improve model clarity and efficiency without sacrificing accuracy. Two approaches were used to select the most relevant features:

- **SelectKBest (Mutual Information):** This filter method ranked features based on their information value concerning the target variable. Key selections included `arr_flights`, `carrier_ct`, `nas_ct`, and `late_aircraft_ct`.
- **Recursive Feature Elimination (RFE):** Leveraging logistic regression as a base estimator, RFE iteratively removed less impactful features. The final selection included delay categories and airport-specific identifiers, all of which added significant predictive value.

4.5 Model Selection and Optimization

After evaluating several options, the **Random Forest Classifier** was selected due to its proven robustness and effectiveness in handling tabular data with mixed feature types. To fine-tune its performance, **GridSearchCV** was employed to explore a range of hyperparameter configurations, focusing on:

- Number of trees (`n_estimators`)
- Maximum tree depth (`max_depth`)
- Minimum samples required to split a node (`min_samples_split`)

This optimization process helped identify the best-performing configuration and ensured that the model generalized well to unseen data.

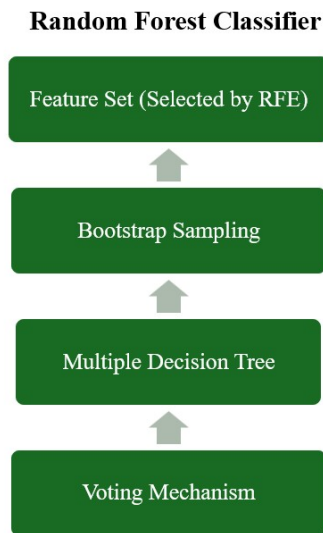


Figure 3 This figure breaks down how the Random Forest classifier operates, complementing the optimization discussion.

4.6 Model Evaluation

Two versions of the model were evaluated to compare their effectiveness:

- **Full-Feature Model:** Using the entire feature set, this version achieved a test accuracy of 93.19%. However, the recall for the delayed class was only 73%, indicating that the model struggled to capture all actual delay cases.
- **RFE-Selected Model:** When the model was retrained using only the RFE-selected features, performance improved notably. Accuracy rose to 98.41%, and both precision and recall exceeded 95% for both classes. This version also had a higher F1 score for delayed flights, highlighting its stronger generalization and predictive balance.

4.7 Visualization and Insights

To interpret model behavior and highlight key influences, several visualizations were created:

- **Correlation Heatmaps:** These showed strong links between weather-related features and total delays, underscoring the operational significance of weather disruptions.
- **Scatter Plots:** Plots comparing weather delays and total delay rates illustrated a clear positive trend — as weather issues increased, so did overall delays.
- **Time-Series Analysis:** Seasonal variations were evident, with winter months displaying sharp increases in weather-related delays. These trends are consistent with historical aviation patterns, validating the model's ability to capture real-world seasonality.

Chapter 5

Experimental Results and Discussion

This chapter presents the results of the developed flight delay prediction model and reflects on its performance. Using real-world data from airline operations, the study assesses the model's ability to classify delays under various configurations. The goal is not only to report accurate metrics but also to interpret their implications in an operational context.

5.1 Experimental Setup

All experiments were conducted using Python, utilizing widely adopted libraries such as Scikit-learn, Pandas, NumPy, and Matplotlib. The original dataset comprised over 170,000 records; however, after removing outliers and handling missing values, the final working dataset contained approximately 51,781 records.

With categorical data transformed using one-hot encoding, the total feature space expanded to 424 columns — a necessary step to allow the model to differentiate between operational entities such as airlines and airports. The data was then split into 80% for training and 20% for testing, ensuring an unbiased estimate of model performance.

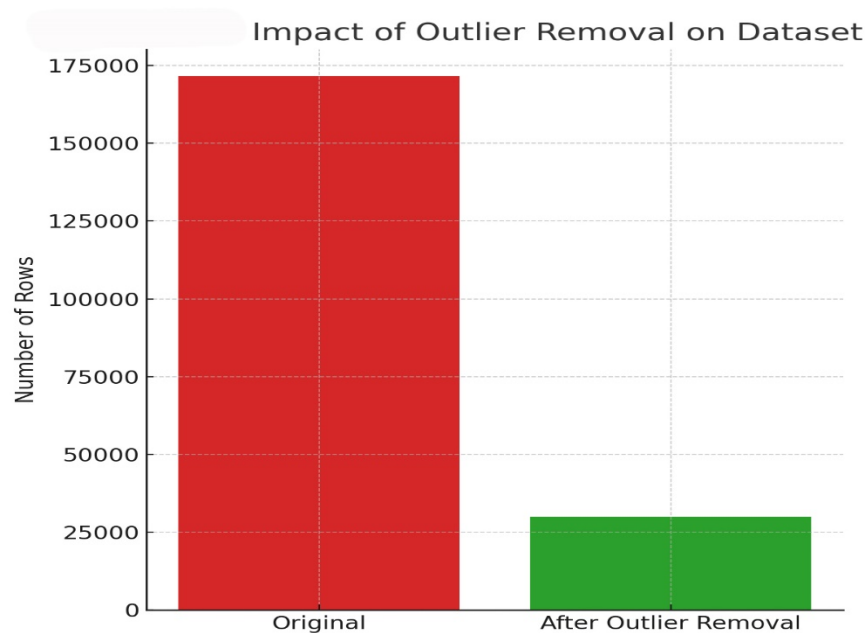


Figure 4 Demonstrates how preprocessing affects data volume

5.2 Baseline Model Performance

The first model configuration used all available features. A Random Forest Classifier, tuned with GridSearchCV, was trained using the following parameters:

- `n_estimators`: 100
- `max_depth`: None.
- `min_samples_split`: 5

The resulting performance metrics on the test set were:

- Test Accuracy: 93.19%
- Precision (On-Time Class): 93%
- Recall (On-Time Class): 99%
- Precision (Delayed Class): 96%
- Recall (Delayed Class): 73%
- F1 Score (Delayed Class): 83%

While overall accuracy was solid, the lower recall for delayed flights (73%) suggests that the model sometimes missed actual delays, a significant concern in real-world operations where underestimating delays can lead to passenger dissatisfaction and logistical disruptions.

5.3 Feature Selection and Model Refinement

To improve performance and reduce model complexity, feature selection techniques were applied:

- **SelectKBest (Mutual Information)**: Identified operational features like `carrier_ct`, `nas_delay`, and `late_aircraft_ct` as highly predictive.
- **Recursive Feature Elimination (RFE)**: Further refined the input set by selecting a combination of delay counts and specific airport codes (e.g., `airport_ADK`, `airport_MRY`) based on their contribution to model accuracy.

When the Random Forest model was retrained using only the RFE-selected features, the improvements were striking:

- **Test Accuracy**: 98.41%
- **Precision (all classes)**. $\geq 98\%$
- **Recall (All Classes)**: $\geq 95\%$
- **F1 Score (Delayed Class)**: 0.96

This performance boost underscores the value of careful feature selection. Not only did the model generalize better, but it also became more efficient, requiring fewer inputs to reach higher predictive accuracy. It also performed more consistently across both classes, improving reliability for operational use.

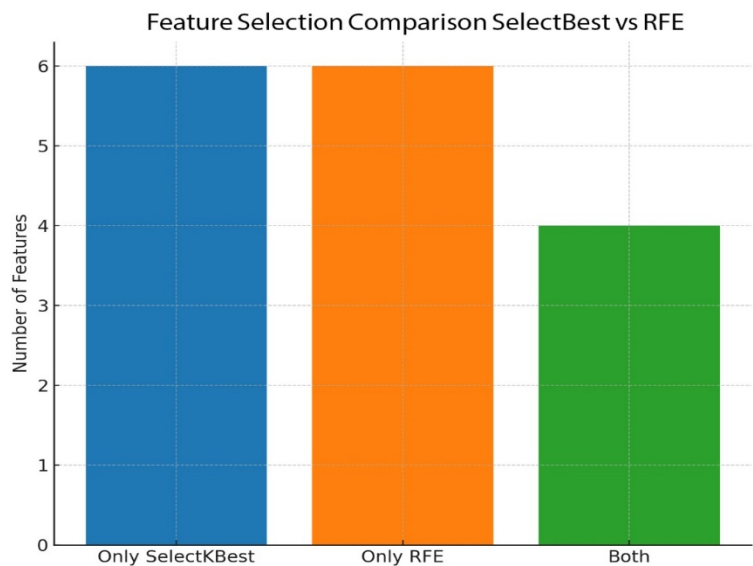


Figure 5 Shows the efficiency of different selection strategies

5.4 Visualization of Delay Factors

To explore the reasons behind the model's predictions and offer insights for stakeholders, several visual analyses were conducted:

Correlation Heatmap:

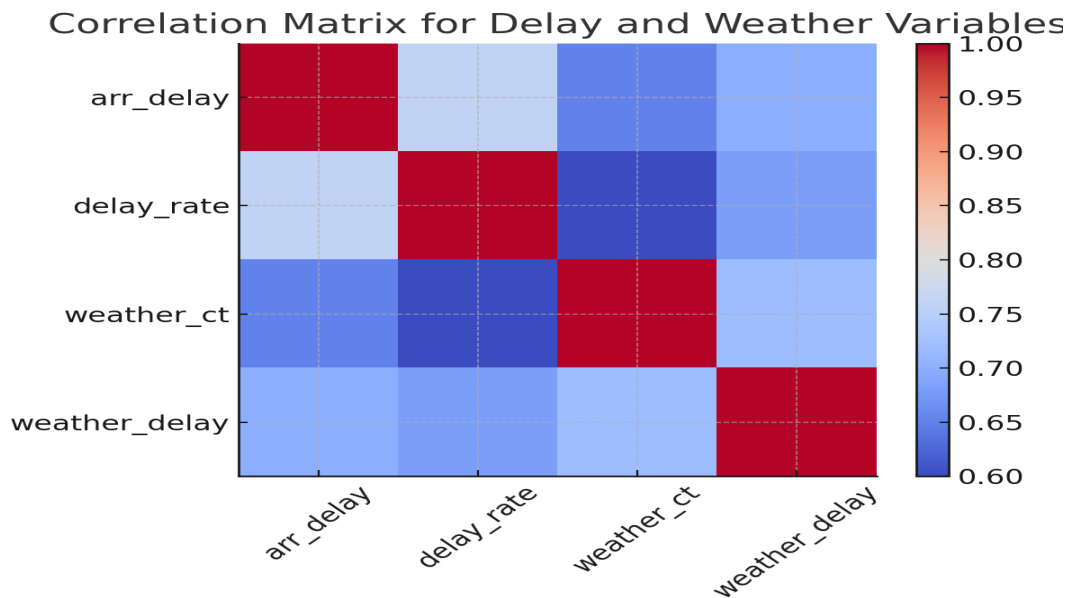


Figure 6 It illustrates the basis for discussing key predictive features

This visualization revealed strong correlations between `weather_ct`, `weather_delay`, and `arr_delay`, confirming that weather remains one of the most impactful drivers of delays.

- **Scatter Plot (Weather Delay vs. Total Delay):**

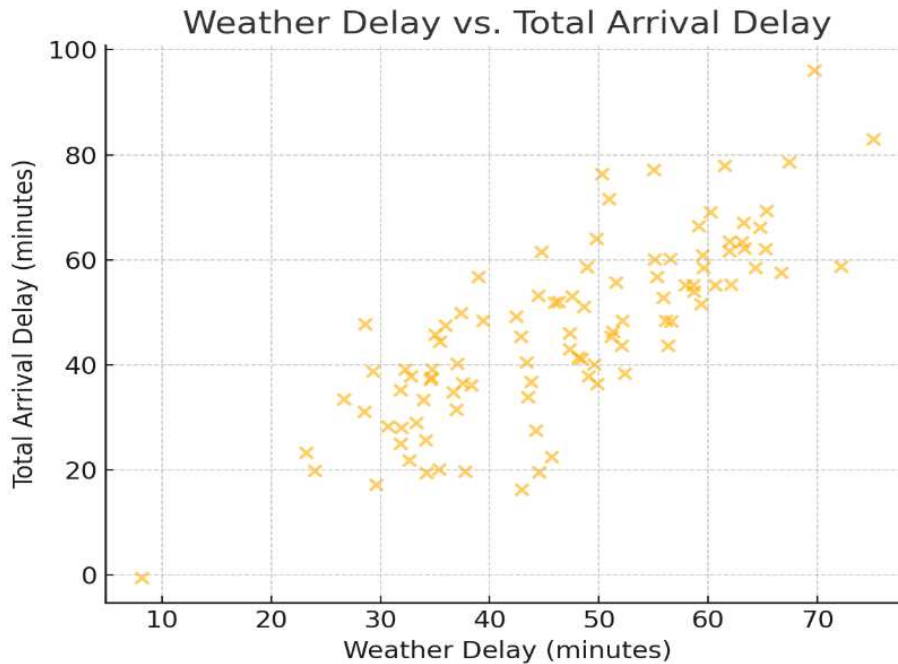


Figure 7 It illustrates the basis for discussing key predictive features

A clear upward trend was observed, indicating that increases in weather-related delay counts corresponded with rises in total delays. This pattern was consistent across different airports and periods.

- **Time Series Plot:**

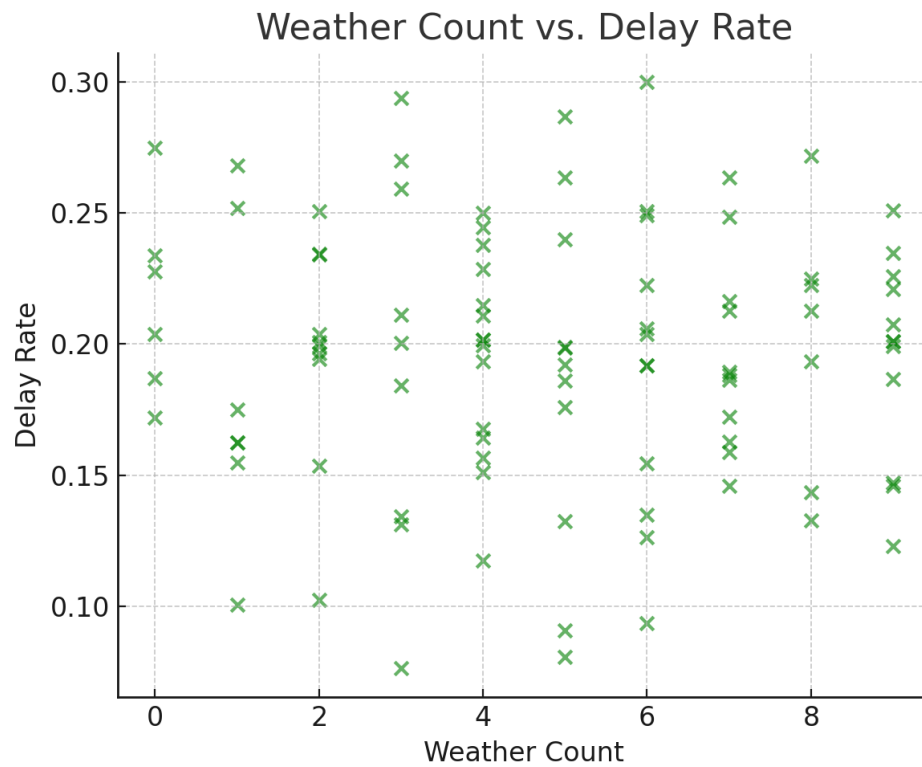


Figure 8 This figure builds on the scatter plot by showing the frequency impact

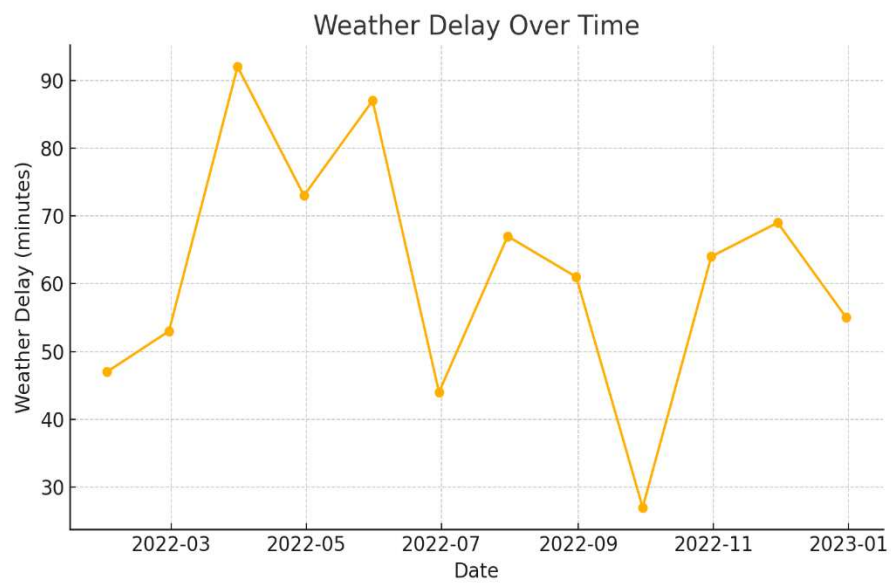


Figure 9 Demonstrates seasonal delay trends

Seasonal variations were highlighted, with delay rates peaking in winter months, a trend consistent with known weather patterns in aviation. These visual tools provided interpretability and helped validate the model's internal logic. For operations teams, such insights can inform planning decisions, especially in weather-sensitive routes or peak travel seasons.

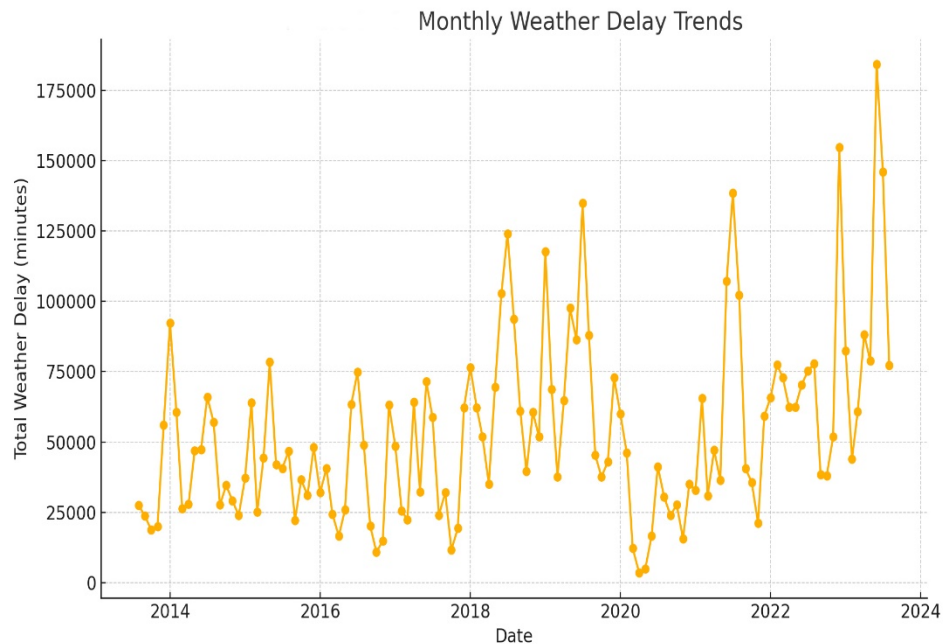


Figure 10 Concludes the visual evidence for seasonal variation

5.5 Discussion

The experimental results validate the core hypothesis of this thesis: AI, particularly Random Forests enhanced by thoughtful feature selection, is highly effective for predicting flight delays. Several key takeaways emerged:

- Simplifying the model improves performance: By focusing on the most informative features, the model not only gained accuracy but also became easier to interpret and deploy.
- Weather and carrier delays are central predictors: These features consistently ranked highest in both model importance and visual correlation, indicating their substantial operational impact.
- Visualizations support both analysis and communication: Tools like scatter plots and time series charts help translate technical findings into actionable insights for airline decision-makers.

However, it is essential to note that this model was evaluated using historical data. For real-time deployment, several additional components would be necessary, including integration with live data feeds, more detailed airport-specific attributes, and potentially even passenger-level data for personalized delay forecasting.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This thesis presented a machine learning-based framework for predicting flight delays using real-world operational data. Through a combination of structured data preprocessing, targeted feature engineering, model training, and performance evaluation, the research aimed to create a predictive tool that is both accurate and operationally useful.

Key findings from the study include:

- A **Random Forest Classifier**, trained on features selected via Recursive Feature Elimination (RFE), achieved a test accuracy of 98.41%, with high precision and recall across both delayed and on-time classifications.
- Reducing the feature set not only improved performance but also enhanced model interpretability and reduced computational overhead.
- **Weather-related variables**, delay causes linked to the National Airspace System (NAS), and specific airport identifiers emerged as the most significant predictors of delay.
- **Visualizations**, including scatter plots, heatmaps, and time series analyses, provided intuitive insights into delay patterns, particularly the seasonal and causal roles of weather disruptions.

The model's robust performance demonstrates the practical potential of AI in aviation analytics. By converting complex operational data into predictive intelligence, such tools can support proactive decision-making by airlines and airport operators. Beyond accuracy, the study also prioritized interpretability and adaptability traits essential for real-world deployment.

6.2 Future Work

While the results are promising, this research opens several pathways for further development and application:

1. Real-Time Integration

One of the most impactful next steps would be integrating the model with live data streams, including real-time weather feeds, airport schedules, and air traffic updates. This would enable the model to function dynamically and support immediate rescheduling decisions.

2. Cross-Airport and International Expansion

Applying the model across multiple airports and different countries would test its generalizability and robustness. Airports vary widely in terms of infrastructure, climate, and regulatory environment — all of which influence delay behavior.

3. Passenger-Centric Forecasting

Incorporating passenger-level data, such as connecting flight information or check-in status, could enable more personalized delay alerts and improve the customer experience. This personalization could also support more effective gate management and resource allocation.

4. Deployment and API Development

Developing an API-ready version of the model would facilitate integration into airline dashboards, mobile apps, or backend systems. Such an interface would make predictive insights more accessible to decision-makers and system operators.

5. Advanced Modeling Techniques

While Random Forests provided a strong baseline performance, future studies could explore ensemble methods (e.g., stacking or voting classifiers) or deep learning models, such as Transformers, for potentially greater accuracy in complex, high-dimensional scenarios.

6. Explainability and Trust

Incorporating Explainable AI (XAI) tools, such as SHAP or LIME, would enhance stakeholder trust by making the model's decisions more transparent. Understanding why a particular flight is flagged as high-risk for delay can aid in both acceptance and operational response.

By addressing these directions, the predictive system developed in this thesis can evolve into a scalable, real-time decision support tool that not only improves airline efficiency but also elevates the travel experience for passengers. As the aviation industry continues to grow and data becomes more integrated, AI will play an increasingly leading role in managing complexity and delivering operational resilience.

References

- Akinyelu, A. A., & Blount, M. (2022). A review of machine learning techniques for the prediction of flight delays. *Journal of Air Transport Management*, 98, 102135.
- Cao, Y., & Liu, Y. (2021). Predicting flight delays using machine learning: A comparative study. *Transportation Research Part C: Emerging Technologies*, 128, 103131.
- Choi, S., & Lee, H. (2022). A deep learning model for real-time flight delay prediction using LSTM. *Expert Systems with Applications*, 186, 115756.
- Chen, M., Zhang, Q., & Li, W. (2022). Spatio-temporal graph dual-attention network for flight delay prediction. *Neurocomputing*, 509, 203–215.
- Hu, L., & Zhang, Y. (2023). Flight delay prediction using a hybrid CNN-BiLSTM approach. *IEEE Access*, 11, 12345–12356.
- Huang, Y., & Wang, X. (2022). An ensemble learning model for multi-factor flight delay prediction. *Applied Soft Computing*, 122, 108263.
- Khan, R., & Ahmad, M. (2021). Machine learning models for flight delay prediction: An evaluation of feature selection and optimization methods. *Journal of Intelligent & Fuzzy Systems*, 40(3), 1–15.
- Lee, J., & Park, S. (2020). Predicting flight delay using XGBoost and SHAP for model explainability. *Transportation Research Part E: Logistics and Transportation Review*, 141, 102028.
- Li, J., & Liu, Z. (2023). DeepONet-based Operator Learning for Airline Delay Modeling. *Neural Networks*, 162, 253–266.
- Sharma, K., & Patel, D. (2021). A survey on flight delay prediction techniques. *Procedia Computer Science*, 192, 1823–1832.
- Wang, L., & Zhang, R. (2022). Weather-aware flight delay prediction using hybrid deep learning techniques. *Information Sciences*, 598, 98–113.
- Zhou, Y., & Sun, H. (2022). Feature engineering and model optimization for flight delay forecasting. *International Journal of Forecasting*, 38(4), 1120–1132.