Predicting Flight Delays with Machine/Deep Learning

Hasan TUNA

Department of Engineering
TOBB University of Economy and Technology
211101017
ht@hasantuna.com

Abstract—Flight delays pose significant challenges to the aviation industry, leading to economic losses and operational disruptions. This paper examines the use of tree-based machine learning algorithms, SVMs, and DNNs to predict whether a flight will be delayed for more than 15 minutes. Using data from U.S. domestic flights, we analyze various factors, including carrier, flight scheduling and flight route to improve delay predictions. Our findings highlight that these models show good enough performance to capture the complexities of flight delays, offering practical insights for enhancing on-time performance and reducing the impact of delays on airlines and passengers. As a result, we have achieved a best accuracy of 0.82 with the Random Forest algorithm and best ROC-AUC score of 0.7 with the DNN algorithm.

I. INTRODUCTION

The aviation industry plays a critical role in global transportation, connecting millions of passengers and goods across vast distances every day. However, flight delays remain a persistent challenge, causing significant economic losses, operational inefficiencies, and widespread dissatisfaction among travelers. The ability to accurately predict flight delays is crucial for airlines, airports, and passengers alike, as it enables better planning, resource allocation, and customer service.

A. Background

There exists some older studies which have used complex feature sets and deep-machine learning algorithms to predict flight delays. However, these studies can be too complex for everyday use.

Traditional flight delay prediction methods have relied on simulations and statistical models, which, while providing valuable insights, often require extensive parameter tuning and are limited by their assumptions. The advent of machine learning has introduced more sophisticated approaches to flight delay prediction, enabling the analysis of vast amounts of data to uncover patterns that were previously difficult to detect.

One prominent approach to flight delay prediction involves using tree-based machine learning algorithms. In a study conducted by Mustafa Kurt (2019), various tree-based algorithms such as decision trees, random forests, and gradient boosting were applied to predict flight departure delays for domestic flights in the U.S. The study utilized data from August 2018, integrating information related to the aircraft, passengers, and cargo with flight data to enhance the model's predictive capabilities. The results highlighted the importance of considering multiple factors, including aircraft age, flight scheduling, and airport operations, in improving the accuracy of delay predictions.

Kurt's study demonstrated that tree-based algorithms could effectively model flight delays, but it also identified areas for improvement. Specifically, the study suggested that adding more features related to flight planning, personnel, and technical processes could enhance the model's predictive power. Moreover, the study underscored the value of hyperparameter tuning and feature engineering in optimizing model performance.

Recent advances have taken flight delay prediction further by incorporating deep learning and spatio-temporal models. For instance, a novel model named CausalNet integrates causal inference with spatio-temporal graph neural networks (STGNNs) to predict flight delays more accurately. CausalNet utilizes Granger causality to construct dynamic causality graphs that capture the complex inter-airport relationships influencing delay propagation. The model's self-corrective mechanism adjusts these graphs in real-time, leading to significant improvements in prediction accuracy.

The combination of these approaches—tree-based algorithms and advanced neural network models—highlights the evolving landscape of flight delay prediction. These methods collectively contribute to a more nuanced understanding of the factors driving delays, offering the potential for more effective mitigation strategies in the aviation industry.

B. Objective

In this study, we focus on the application of various machine learning algorithms, including tree-based models, Support Vector Machines (SVMs), and Deep Neural Networks (DNNs), to predict whether a flight will be delayed for more than 15 minutes. We utilize a comprehensive dataset from U.S. domestic flights, incorporating key features such as carrier information, flight scheduling, and route details to improve the accuracy of delay predictions.

Our analysis demonstrates the effectiveness of these machine learning models in capturing the complexities of flight delays. Specifically, the Random Forest algorithm achieved the highest accuracy of 0.82, while the DNN model yielded the best ROC-AUC score of 0.7 (scored on Kaggle). These results

suggest that machine learning can significantly enhance on-time performance and mitigate the impact of delays on both airlines and passengers.

C. Structure of the Paper

The remainder of this paper is organized as follows: Methodology provides an overview of the related work in flight delay prediction. Section 3 describes the dataset and features used in the study. Section 4 details the machine learning methods applied, while Section 5 presents the results and discusses their implications. Finally, Section 6 concludes the paper with recommendations for future research.

II. Methodology

A. Feature Engineering

Almost all of the models need numerical data to work. So, we need to convert the categorical features into numerical features. We can use one-hot encoding or label encoding for this purpose. We have tried both and we have seen that one-hot encoding creates too many features and it makes the model slower. So, we have decided to use label encoding.

After applying label encoding, the categorical features will be converted into numerical features that can be used by the machine learning models. This step is essential for ensuring that the models can process the data effectively and make accurate predictions.

Once the categorical features are encoded, we have normalized the numerical features to ensure that they are on the same scale. Normalization helps prevent features with larger values from dominating the model and ensures that all features contribute equally to the prediction.

After all the conversions are done, we have looked at the summary of the data and saw that the dataset is highly biased, with %80 of the flights are on time. This situation means that no-learning models can achieve %80 accuracy. So, we have decided to use ROC-AUC score as the main metric for evaluating the models.

B. Model Selection

We have used the following machine learning algorithms to predict flight delays:
- Decision Trees
- Random Forest
- Support Vector Machines (SVM)
- Deep Neural Networks (DNN)

We have tried different hyperparameters for each algorithm and we have selected the best performing hyperparameters for each algorithm. We have used the scikit-learn library for Decision Trees, Random Forest, and SVMs, and the TensorFlow library for DNNs.

There are 8 different Decision Tree models, each having different max_depth values between 8 and 64, 5 Random Forest models, each having different max_feature values between 1 and 5, 1 SVM model with a linear kernel, 1 DNN, 1 Embedded DNN (Using Embedding Layers for feature transformation), 1 Modified Random Forest (Using only 3 most effective features).

C. Model Evaluation

We have used the ROC-AUC score as the main metric for evaluating the models. We have also used the accuracy score to see the overall performance of the models. We have used the confusion matrix to see the performance of the models in more detail.

III. Results

A. Decision Trees

We have trained 8 different Decision Tree models with different max_depth values between 8 and 64. The best performing model has a max_depth value of 8 and achieved an accuracy of 0.81 and an ROC-AUC score of 0.57. Increasing the max_depth value decreased the performance of the model. This is probably due to overfitting.

B. Random Forest

We have trained 5 different Random Forest models with different max_feature values between 1 and 5. The best performing model has a max_feature value of 5 and achieved an accuracy of 0.82 and an ROC-AUC score of 0.56.

And the Modified Random Forest model achieved an accuracy of 0.81 and an ROC-AUC score of 0.60.

Using only the most effective features increased the performance of the model.

C. Support Vector Machines

We have trained an SVM model with a linear kernel. The model achieved an accuracy of 0.81 and an ROC-AUC score of 0.5. This is basically no-learning. Which makes sense because of the dataset is not linearly separable.

D. DNN

We have trained a DNN model using keras. The model achieved an accuracy of 0.81 and an ROC-AUC score of 0.63 (on kaggle).

And the Embedded DNN model achieved an accuracy of 0.81 and an ROC-AUC score of 0.7 (on kaggle). Using embedding layers for feature transformation increased the performance of the model. And this is the best performing model in this study.

IV. Discussion

The results of this study demonstrate the effectiveness of machine learning algorithms in predicting flight delays. The Random Forest algorithm achieved the highest accuracy of 0.82, while the DNN model yielded the best ROC-AUC score of 0.7. These findings suggest that it is possible to create a simpler model to to predict flight delays good enough.

Sadly, the accuracy is not beyond no-learning models. This is probably due to the dataset is not highly biased. We can try to use more complex models to increase the performance of the models. However, this study is focused on simpler models. Even with the bias, we were able to achieve 0.7 ROC-AUC which is a good result.

To improve the performance of the models, we can consider the following strategies:

- Feature Engineering: We can explore additional features related to flight operations, weather conditions, and airport congestion to enhance the predictive power of the models.
- Hyperparameter Tuning: We can further optimize the hyperparameters of the models to improve their performance and generalization capabilities.
- Ensemble Methods: We can combine the predictions of multiple models to create a more robust and accurate prediction system. A network of random forests and DNNs can be used for this purpose. However we have not tried this in this study.
- Bias-Resolving: We can use techniques such as SMOTE to balance the dataset and reduce bias in the models. This can help improve the performance of the models on imbalanced datasets.