Enhancing Aviation Efficiency through Big Data and Machine Learning for Flight Delay Prediction

Gerasimos Vonitsanos[1], Ioannis Gounaridis[1], Andreas Kanavos[2] and Phivos Mylonas[3]

[1] Computer Engineering and Informatics Department
University of Patras, Patras, Greece
mvonitsanos@ceid.upatras.gr, igounaridis@upatras.gr
[2] Department of Informatics,
Ionian University, Corfu, Greece
akanavos@ionio.gr
[3] Department of Informatics and Computer Engineering
University of West Attica, Athens, Greece
mylonasf@uniwa.gr

Abstract. Flight delays pose significant challenges to the aviation industry, leading to increased operational costs and passenger dissatisfaction. This paper explores the use of machine learning (ML) and big data analytics to enhance the accuracy and efficiency of flight delay predictions. Utilizing data from the Federal Aviation Administration (FAA) covering the period from 2018 to 2022, we analyze critical factors influencing delays and develop predictive models employing techniques such as Random Forest, Gradient Boosting Machines, Decision Trees, and k-Nearest Neighbors. Our analysis demonstrates that these ML techniques significantly outperform traditional models, improving the accuracy of delay predictions and thereby supporting airlines in optimizing operational efficiency and enhancing passenger satisfaction. The paper also discusses the practical implementation of these findings in real-time airline operations and outlines future research directions to further improve predictive accuracy.

Keywords: Flight Delays · Machine Learning · Predictive Modeling · Big Data Analytics · Random Forest · Gradient Boosting Machines · k-Nearest Neighbors · Operational Efficiency

1 Introduction

Flight delays are a pervasive challenge in the aviation industry, influenced by a myriad of factors that compromise the efficiency and reliability of air travel [24]. Weather conditions, such as fog, thunderstorms, and heavy snow, often pose significant disruptions, as they can severely impact visibility and flight schedules, necessitating delays or outright cancellations. Beyond meteorological issues,

technical faults in aircraft, including engine failures or electronics malfunctions, further contribute to operational delays [19]. These technical disruptions require extensive checks and repairs, often leading to prolonged ground times. Additionally, operational challenges such as inefficient boarding processes, baggage handling delays, and logistical complications with connecting flights compound the problem. The cumulative effect of these issues not only escalates operational costs for airlines but also diminishes passenger satisfaction due to increased travel times, missed connections, and the ensuing stress and inconvenience [9].

In response to these challenges, the integration of advanced technologies, particularly artificial intelligence (AI), robotics, and digital systems, is reshaping how the aviation industry addresses flight delays. AI and machine learning are at the forefront of this technological revolution, offering sophisticated tools that enhance decision-making and operational efficiency [6,16]. For instance, machine learning algorithms can now predict weather patterns more accurately, allowing for better preparation and adjustment of flight schedules to minimize weather-related delays [29]. Similarly, robotics and automated systems are being deployed to streamline aircraft maintenance and baggage handling, thereby reducing the likelihood of delays related to technical issues or ground operations. The potential of these technologies extends into nearly every aspect of aviation operations, promising significant reductions in delay frequencies and durations [8].

Our research makes a substantial contribution to this field by developing advanced machine learning models that harness vast datasets to predict and mitigate flight delays with unprecedented accuracy. We focus on constructing predictive models that analyze historical flight data, weather reports, and operational metrics to identify the most significant predictors of delays. By integrating techniques such as Random Forest, Gradient Boosting Machines, and k-Nearest Neighbors, our models achieve a fine balance between accuracy and computational efficiency. Furthermore, we explore the application of these models in real-time settings, providing airlines with actionable insights that can be immediately implemented to avoid potential delays. Additionally, our study extends into the realm of enhancing passenger experiences by using machine learning to offer personalized travel recommendations, tailored in-flight services, and efficient loyalty programs. These improvements not only bolster customer satisfaction but also fortify brand loyalty and open new revenue streams for airlines.

Through detailed analysis and innovative applications of machine learning, our paper illustrates how data-driven technologies can transform the aviation industry by mitigating the impact of flight delays and refining the overall travel experience. By presenting novel methodologies and demonstrating their effectiveness in real-world scenarios, we aim to provide a blueprint for airlines seeking to improve operational efficiency and customer service through technological innovation.

The remainder of the paper is structured as follows: Section 2 reviews the related work, highlighting previous studies and advancements in the application

of machine learning techniques to predict flight delays. Section 3 details the foundational methodologies employed, including the specific machine learning algorithms and data handling techniques used. Section 4 describes the practical implementation of the models, from data preprocessing to feature selection. Section 5 evaluates the performance of the implemented models against various metrics to assess their effectiveness in predicting flight delays accurately. Finally, Section 6 concludes the paper with a summary of findings and discusses potential directions for future research to further enhance the predictive capabilities and operational applications.

2 Related Work

The aviation sector, a crucial component of the economy and modern life, is significantly affected by flight delays, impacting both passengers and airlines [2]. It has been identified that big data analysis can address this issue in a sophisticated manner. Through the use of modern machine learning techniques, it is possible for airlines to analyze data from previous flights to predict potential delays. Such predictions enable proactive measures to be taken, such as informing passengers or avoiding problems before they escalate. Furthermore, the analysis of large volumes of data allows for a better understanding of the factors causing delays, aiding airlines in improving their efficiency and reducing financial losses. Generally, the benefits of big data analysis for airlines include the prevention of delays, enhanced operational efficiency, and minimized financial losses.

The synthesis of multiple data sources is recognized as a critical element in accurately predicting flight delays. Information collected from various sources, including weather stations, aircraft traffic control systems, and historical flight data, facilitates an integrated approach [3]. Advanced algorithms are employed to analyze these combined data sources, enhancing the accuracy of delay predictions by considering multidimensional parameters and dynamics [30].

Moreover, the application of big data and machine learning technologies in the aviation sector is paving the way for advanced solutions beyond simple delay prediction. These technologies are instrumental in improving aircraft traffic management, enhancing flight safety, and improving passenger experiences. This capability underscores the importance of integrating these technologies into various aspects of the aviation industry, highlighting their multidimensional value [1].

The continuous evolution and improvement of machine learning methods are providing the potential for robust and flexible prediction models. Such models are capable of quickly adapting to new conditions and changes, such as unpredictable weather conditions or operational changes at airports [11]. This flexibility is deemed critical for effective delay forecasting in the dynamic and complex aviation sector [22].

The examination of large volumes of data allows for addressing the issue of flight delays in a more advanced manner. By employing modern machine learning techniques, such as Random Forest, companies are enabled to analyze

large datasets to predict potential delays and take proactive measures. This approach facilitates informing passengers or preparing for controls that can prevent problems at the last minute [31]. However, the challenge is not only limited to the selection and implementation of appropriate algorithms but also extends to the management and assurance of data quality. Data cleanliness, accuracy, and up-to-dateness are crucial aspects that directly impact the outcomes of the algorithms. Ensuring high-quality data and keeping it continuously updated are essential for producing reliable and accurate forecasts, making data science an integral part of the flight delay prediction process [5].

Additionally, the study by [10] explores the use of deep neural networks for predicting aviation demand in time series data, applying several models to determine the most effective strategy. In another recent development, a system combining LSTM and CNN technologies demonstrated notable improvements in classification tasks [21]. This approach notably reduced execution times by 30% to 42%, thereby underscoring the significant advantages of LSTM neural networks in specific applications. In another field, authors in [15,26] focused on Twitter sentiment analysis for the classification of user sentiments in tweets about COVID-19 on Twitter and implemented sentiment analysis using seven different deep learning models based on LSTM neural networks.

The integration of multiple data sources is reiterated as a key element in the accurate prediction of flight delays. Data from various sources, such as weather stations, aircraft traffic control systems, and historical flight data, allows for a more holistic approach. The combined analysis of these data through advanced algorithms aids in more accurately predicting delays, taking into account multidimensional parameters and dynamics [30]. Big data analysis facilitates the identification of various factors that can cause delays, such as economic conditions, human interference, or technical problems [12,27]. Through this analytical process, airlines can enhance their efficiency, reduce financial losses, and provide a better experience to their customers. Any problem or phenomenon can be addressed more effectively through a scientific approach and the utilization of available data, thus achieving a more thorough understanding and interpretation of the issue [4].

3 Methodology Foundations

The methodologies employed in this research are foundational to enhancing the predictive accuracy of flight delay models. By deploying a range of advanced machine learning algorithms and predictive modeling techniques, this section aims to delineate the approaches used to harness and analyze vast datasets. These methods not only predict delays more effectively but also help in understanding the complex dynamics of aviation operations. This dual focus on Predictive Modeling and Machine Learning Models forms the core of our approach, ensuring both the reliability and applicability of our findings in real-world aviation scenarios.

3.1 Predictive Modeling Techniques

In aviation research, a consistent and systematic approach to analyzing and predicting air traffic delays is essential. This overview of Predictive Modeling Techniques serves as an introduction to the principal methods utilized in this field. Selecting suitable algorithms for delay prediction is crucial, given that the effectiveness of an algorithm can vary with the nature and volume of the data. By developing and training models, our aim is to optimize prediction accuracy. Evaluating and validating these models is imperative to ensure their accuracy and reliability.

The model was trained to predict airline flight delays using variables such as departure time, destination, weather conditions, and other relevant factors. Predictive modeling has proven invaluable across various domains, including finance, health, industry, and science. Specifically, in aviation, it aids airlines in enhancing their performance and delivering superior service to passengers.

One significant advantage of Predictive Modeling is its capacity to analyze historical data and forecast future trends or events. Employing machine learning and statistical analysis algorithms, predictive models can identify data patterns and project how these patterns will evolve. Developing such models is crucial for making real-time decisions, enhancing system performance, and minimizing risks [14].

3.2 Advanced Machine Learning Algorithms

A pivotal aspect of applying machine learning is the selection of appropriate algorithms, ensuring that the predictions generated are accurate and dependable. After evaluating the dataset characteristics and the specific requirements of this research, the chosen algorithms include Decision Trees, Random Forests, Gradient Boosting Machines, k-Nearest Neighbours, and Support Vector Machines.

The Decision Trees algorithm is a supervised method used to discern relationships and patterns between input and target features, represented in a model. The strength of Decision Trees lies in their simplicity and interpretability; they easily handle qualitative predictors without the need for dummy variables and are intuitively understandable to humans. Whether it is a Classification Tree or Regression Tree depends on the nature of the input characteristics—categorical or continuous, respectively [18].

Random Forest is a supervised classification technique that enhances prediction accuracy by employing multiple classifiers. This ensemble approach creates a forest of decision trees where each tree is trained on a random subset of the data and features, making the final model more robust against overfitting. Random Forests are particularly effective in managing extensive datasets, where individual classifiers might falter, thereby complementing the decision tree approach with improved stability and accuracy [23].

Gradient Boosting Machines (GBM) algorithm stands out in dealing with complex data sets and non-linear relationships. It incrementally improves models based on the prediction errors of preceding iterations, focusing on points of

failure. By amalgamating numerous weak models, such as decision trees, GBM assembles a robust predictive model, applicable in various scenarios like fraud detection and market forecasting. The adaptability of GBM to different types of data and its ability to handle heterogeneous features make it a powerful tool for predictive modeling [17].

The k-Nearest Neighbours (kNN) method is renowned for its simplicity and efficacy, utilizing a non-parametric approach to classification. This algorithm classifies new instances based on a majority vote of the k nearest neighbors, where k is a user-defined constant, and neighbors are taken from a set of objects for which the correct classification is known. This method is highly adaptable and effective in scenarios where the decision boundary is irregular [20].

Lastly, Support Vector Machines (SVMs) are utilized for their superior performance and generalization capabilities compared to other classifiers. SVMs excel in high-dimensional spaces and are particularly effective in cases where the number of dimensions exceeds the number of samples. Their ability to use a kernel trick to handle linear and non-linear separation makes them versatile for a variety of classification problems. SVMs are adept at solving problems across three categories—linearly separable, linearly nonseparable, and nonlinearly separable—by finding the optimal hyperplane for data classification [7].

4 Implementation

4.1 Data Preparation and Preprocessing

The data used in this study were sourced from the widely recognized Kaggle platform, specifically from the "Marketing Carrier On-Time Performance (Beginning January 2018)" data table of the "On-Time" database within the TranStats data library. Concretely, covering airline flights from 2018 to 2022, this dataset is invaluable for its detailed information on flights, weather conditions, airline operators, and airports. Such comprehensive data is essential for dissecting the complex factors that influence flight delays and for crafting predictive models with high accuracy.

Data preprocessing is a critical step that significantly influences the performance and reliability of machine learning algorithms. This process involves several key tasks: handling missing values to prevent biases, normalizing features to bring data onto a similar scale, reducing dimensionality to focus on relevant information, and removing noise to improve model accuracy. These steps are fundamental in transforming raw data into a refined format suitable for analysis, ensuring that subsequent modeling is based on clean and meaningful data.

4.2 Feature Selection and Justification

Effective feature selection is paramount in machine learning to enhance model performance and ensure computational efficiency. This process entails choosing the most impactful features from a dataset, which can dramatically reduce the

complexity of the model while improving its predictive power. Feature selection techniques are categorized into three main types: filters, embedded methods, and wrapper methods. Filter methods evaluate features based on statistics independent of the model; embedded methods incorporate feature selection as part of the model training process, often using regularization; wrapper methods evaluate subsets of features by actually training models on them and seeing their effect on performance [13].

In this paper, the selection of features was guided by several criteria aimed at maximizing predictive accuracy and model robustness. Features were selected based on their strong correlation with the target variable, minimal redundancy with other features, and their ease of collection in real-world scenarios. This rigorous selection process helps prevent overfitting, enhances model generalizability, and leads to more efficient training and execution [25]. Furthermore, by systematically identifying and employing the most impactful features, the models are better equipped to handle vast datasets effectively, ultimately improving the speed and accuracy of predictions in operational settings.

Following an in-depth analysis, certain features were identified as having significant relationships with flight delays. These features, listed in the following Table 1, represent a mix of temporal, operational, and environmental variables, each contributing uniquely to the model's ability to forecast delays in an accurate way:

Table 1. Selected Features for Flight Delay Prediction

Field
Year
Month
Day of the Week
Day of Flight
Unique Airline Codes
Name of the Airline
Original Airport ID
Name of Destination City
Airport Destination ID
Flight Delay
Flight Cancellation
Change of Flight Path

The fields not directly related to flight delays were deemed non-essential for the analysis and were excluded. This decision was based on their negligible contribution to the understanding of flight delays, ensuring that the model remains focused on the most predictive features.

This approach not only streamlines the feature set but also boosts the model's efficiency and effectiveness, essential for deploying these models in real-time environments where quick and accurate predictions are necessary.

5 Evaluation

Evaluating the effectiveness of predictive models is crucial to ensure they are reliable and accurate for their intended real-world applications. This section delves deeper into the various aspects of model evaluation, exploring the complexity of the data, the effectiveness of different machine learning models, and the implications of their performance metrics.

5.1 Data Complexity and Evaluation Metrics

The data concerning air traffic delays encompasses a vast array of variables, including temporal elements like the date and time of flights, spatial details such as airport locations and routes, and statistical distributions such as the frequency and magnitude of delays. Given this complexity, choosing the right metrics for evaluation is paramount.

The primary metric used for assessing the models is the Root Mean Squared Error (RMSE), which measures the average magnitude of the prediction error. RMSE is particularly valued in regression problems because it highlights larger errors, providing insight into how well a model can predict significant deviations from the norm. This is crucial in applications like flight delay predictions, where accurately forecasting significant delays can prevent large-scale disruptions [32].

5.2 Performance Evaluation of Machine Learning Models

To rigorously evaluate the performance of each machine learning model, we considered several algorithms: Random Forest, Gradient Boosting Machines (GBM), Decision Trees, and k-Nearest Neighbors (KNN). Each model's performance was scrutinized using not only RMSE but also other important metrics such as accuracy, recall, and the F1 score, which collectively offer a fuller picture of model capabilities.

Before discussing the results, it is essential to view the summarized outcomes illustrated in Figure 1, which compares the RMSE values for different models across several years.

This figure clearly indicates that the GBM model consistently achieves lower RMSE scores across the years compared to other models, suggesting its effectiveness in handling the nuances and complexities of flight delay data. The Decision Trees and KNN models show higher variability and generally higher RMSE values, which might indicate issues with overfitting or inadequate handling of the dataset's diverse features.

Furthermore, a detailed comparison of the yearly performance of each model is presented in the table below. Table 2 allows us to observe the consistency and variability in the performance of each algorithm over time.

The table corroborates the graphical analysis, highlighting that while GBM and RF are more consistent, DT and KNN show significant fluctuations in performance year by year. This suggests that while GBM and RF can be relied upon for more stable performance across varying conditions, DT and KNN may

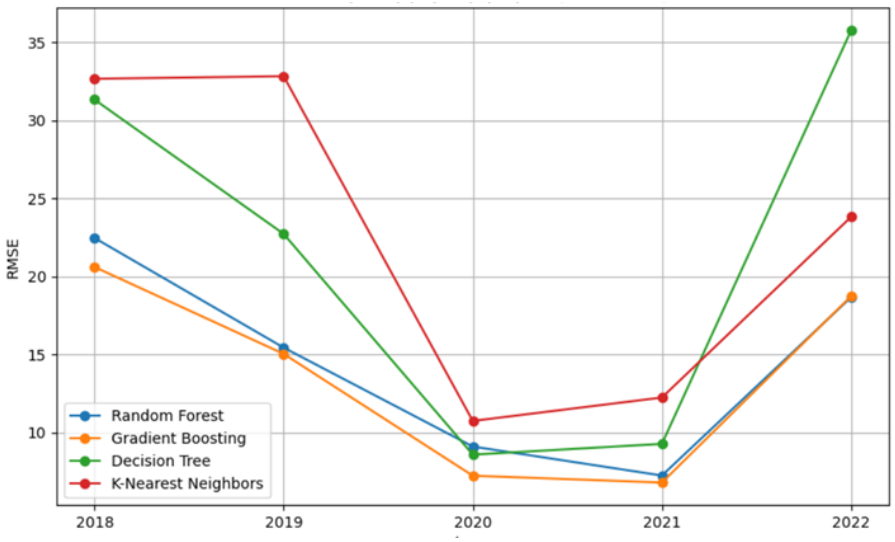

Fig. 1. Comparison of RMSE values for the years 2018-2022 across different machine learning models, illustrating their performance in predicting flight delays

Table 2. Comparison of machine learning algorithms over years 2018-2022, showing detailed RMSE values to underscore differences in model performance

Year	GBM	RF	DT	KNN
2018	20.60	22.46	31.33	32.66
2019	15.03	15.43	22.72	32.83
2020	7.22	9.09	8.57	10.73
2021	6.78	7.23	9.26	12.23
2022	18.73	18.66	35.76	23.81

require further tuning or might be less suitable for this specific application due to their sensitivity to the data's diverse conditions.

5.3 Discussion

The analysis provided by Figure 1 and Table 2 reveals distinct trends in model performance. The consistently low RMSE values of the GBM algorithm across the evaluated years indicate not only high accuracy but also robustness under varying conditions and datasets. This suggests that GBM is particularly effective for predicting flight delays with high reliability.

Conversely, the k-Nearest Neighbors (KNN) method shows the highest RMSE, particularly in the earlier years, suggesting limitations in its ability to handle the complexity and variability of the flight delay data effectively. The Random Forest and Decision Trees models show variability in their performance, with

generally good results that highlight their strengths in handling nonlinear data relationships.

These findings emphasize the need for careful selection of predictive models based on specific data characteristics and the particular requirements of the predictive tasks. The insights provided here offer a foundation for future research and application enhancements, focusing on optimizing model selection and tuning to improve prediction accuracy in practical settings.

6 Conclusions and Future Work

Airport delays pose a significant challenge within the aviation industry, affecting not only passenger satisfaction but also the operational efficiency of airlines. Such delays impact travel schedules, overall travel time, and the general perception of the airline industry. With increasing airport traffic and limited resources, there is a pressing need for innovative solutions that can predict and manage these delays effectively.

This research has demonstrated the potential of machine learning models to predict flight delays with high accuracy. Leveraging complex datasets, the models developed provide airlines with powerful tools to enhance their operational strategies. The effectiveness of these techniques suggests substantial opportunities for practical applications in flight management and planning.

Looking forward, future research could benefit from the integration of additional variables such as real-time weather conditions, crew scheduling, and passenger data, which could further refine the accuracy of delay predictions. The methodologies applied in this study hold potential for adaptation across other public transportation systems, such as railways, shipping, and buses, each introducing unique challenges and variables that require tailored predictive models [28].

Another promising direction involves the development of real-time prediction applications that provide passengers with timely alerts about potential delays. Such applications would need to dynamically update predictions as new data becomes available, enhancing the traveler experience. Additionally, to ensure the continued relevance and effectiveness of these models, regular updates to the datasets and continuous refinement of the algorithms are essential. Collaborating with airlines would not only provide access to richer datasets but also facilitate the practical implementation of research findings and allow for iterative improvements based on direct feedback.

The ongoing evolution of the aviation sector, coupled with advances in machine learning and data analytics, underscores the importance of continuous research and innovation. By improving predictive accuracy and operational responsiveness, airlines can significantly enhance passenger satisfaction and achieve greater operational efficiency. Future studies should also consider the unpredictable elements that can affect transportation, such as sudden weather changes or geopolitical events, by integrating these factors into more resilient predictive models.

In conclusion, this research provides a foundation for future advancements in the management of flight delays and can serve as a blueprint for addressing similar challenges in other sectors of public transportation. Continuous improvement and adaptation to new data and technologies remain key to advancing this field.

References

1. Alla, H., Moumoun, L., Balouki, Y.: A multilayer perceptron neural network with selective-data training for flight arrival delay prediction. Scientific Programming **2021**, 5558918:1–5558918:12 (2021)
2. Baker, D., Merkert, R., Kamruzzaman, M.: Regional aviation and economic growth: cointegration and causality analysis in australia. Journal of Transport Geography **43**, 140–150 (2015)
3. Bao, Y., Xiong, T., Hu, Z.: Forecasting air passenger traffic by support vector machines with ensemble empirical mode decomposition and slope-based method. Discrete Dynamics in Nature and Society **2012** (2012)
4. Cai, K., Li, Y., Fang, Y., Zhu, Y.: A deep learning approach for flight delay prediction through time-evolving graphs. IEEE Transactions on Intelligent Transportation Systems **23**(8), 11397–11407 (2022)
5. Carvalho, L., Sternberg, A., Goncalves, L.M., Cruz, A.B., Soares, J.A., Brandão, D., Carvalho, D., Ogasawara, E.: On the relevance of data science for flight delay research: A systematic review. Transport Reviews **41**(4), 499–528 (2021)
6. Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z., Zhao, D.: Flight delay prediction based on aviation big data and machine learning. IEEE Transactions on Vehicular Technology **69**(1), 140–150 (2020)
7. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Machine Learning **46**(1-3), 389–422 (2002)
8. Jiang, Y., Liu, Y., Liu, D., Song, H.: Applying machine learning to aviation big data for flight delay prediction. In: IEEE International Symposium on Dependable, Autonomic and Secure Computing (DASC). pp. 665–672 (2020)
9. Jin, F., Li, Y., Sun, S., Li, H.: Forecasting air passenger demand with a new hybrid ensemble approach. Journal of Air Transport Management **83**, 101744 (2020)
10. Kanavos, A., Kounelis, F., Iliadis, L., Makris, C.: Deep learning models for forecasting aviation demand time series. Neural Computing and Applications **33**(23), 16329–16343 (2021)
11. Kanavos, A., Trigka, M., Dritsas, E., Vonitsanos, G., Mylonas, P.: A regularization-based big data framework for winter precipitation forecasting on streaming data. Electronics **10**(16), 1872 (2021)
12. Karamitsos, I., Papadaki, M., Al-Hussaeni, K., Kanavos, A.: Transforming airport security: Enhancing efficiency through blockchain smart contracts. Electronics **12**(21), 4492 (2023)
13. Kumar, V., Minz, S.: Feature selection: A literature review. Smart Computing Review **4**(3), 211–229 (2014)
14. Lantz, B.: Machine Learning with R: Expert Techniques for Predictive Modeling. Packt Publishing (2019)
15. Lyras, A., Vernikou, S., Kanavos, A., Sioutas, S., Mylonas, P.: Modeling credibility in social big data using LSTM neural networks. In: 17th International Conference on Web Information Systems and Technologies (WEBIST). pp. 599–606 (2021)

16. Manna, S., Biswas, S., Kundu, R., Rakshit, S., Gupta, P., Barman, S.: A statistical approach to predict flight delay using gradient boosted decision tree. In: International Conference on Computational Intelligence in Data Science (ICCIDS). pp. 1–5. IEEE (2017)
17. Natekin, A., Knoll, A.C.: Gradient boosting machines, a tutorial. Frontiers Neurorobotics **7**, 21 (2013)
18. Ntaliakouras, N., Vonitsanos, G., Kanavos, A., Dritsas, E.: An apache spark methodology for forecasting tourism demand in greece. In: 10th International Conference on Information, Intelligence, Systems and Applications (IISA). pp. 1–5. IEEE (2019)
19. Qu, J., Zhao, T., Ye, M., Li, J., Liu, C.: Flight delay prediction using deep convolutional neural network based on fusion of meteorological data. Neural Processing Letters **52**(2), 1461–1484 (2020)
20. Sarker, I.H.: Machine learning: Algorithms, real-world applications and research directions. SN Computer Science **2**(3), 160 (2021)
21. Savvopoulos, A., Kanavos, A., Mylonas, P., Sioutas, S.: LSTM accelerator for convolutional object identification. Algorithms **11**(10), 157 (2018)
22. Schösser, D., Schönberger, J.: On the performance of machine learning based flight delay prediction–investigating the impact of short-term features. Promet-Traffic & Transportation **34**(6), 825–838 (2022)
23. Shaik, A.B., Srinivasan, S.: A brief survey on random forest ensembles in classification model. In: International Conference on Innovative Computing and Communications (ICICC). pp. 253–260. Springer (2019)
24. Sternberg, A., de Abreu Soares, J., Carvalho, D., Ogasawara, E.S.: A review on flight delay prediction. CoRR **abs/1703.06118** (2017)
25. Venkatesh, B., Anuradha, J.: A review of feature selection and its methods. Cybernetics and information technologies **19**(1), 3–26 (2019)
26. Vernikou, S., Lyras, A., Kanavos, A.: Multiclass sentiment analysis on covid-19-related tweets using deep learning models. Neural Computing and Applications **34**(22), 19615–19627 (2022)
27. Vonitsanos, G., Kanavos, A., Mylonas, P.: Decoding gender on social networks: An in-depth analysis of language in online discussions using natural language processing and machine learning. In: IEEE International Conference on Big Data. pp. 4618–4625 (2023)
28. Vonitsanos, G., Panagiotakopoulos, T., Kanavos, A., Kameas, A.: An apache spark framework for iot-enabled waste management in smart cities. In: 12th Hellenic Conference on Artificial Intelligence. pp. 1–7 (2022)
29. Vonitsanos, G., Panagiotakopoulos, T., Kanavos, A., Tsakalidis, A.: Forecasting air flight delays and enabling smart airport services in apache spark. In: Artificial Intelligence Applications and Innovations. Springer International Publishing (2021)
30. Yazdi, M.F., Kamel, S.R., Chabok, S.J.S.M., Kheirabadi, M.: Flight delay prediction based on deep learning and levenberg-marquart algorithm. Journal of Big Data **7**(1), 106 (2020)
31. Yi, J., Zhang, H., Liu, H., Zhong, G., Li, G.: Flight delay classification prediction based on stacking algorithm. Journal of Advanced Transportation **2021**(1), 4292778 (2021)
32. Zhou, J., Gandomi, A.H., Chen, F., Holzinger, A.: Evaluating the quality of machine learning explanations: A survey on methods and metrics. Electronics **10**(5), 593 (2021)