
ISSN: 2321-2152
IJMECE
International Journal of modern
electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 10, Issue.4 Dec 2022

FLIGHT DELAY PREDICTION BASED ON AVIATION BIG DATA AND MACHINE LEARNING

Mrs T. Shruthi[1],D. Rajitha[2],A. Shruthi[3],P. Manichandana[4],R. Himabindhu[5],V. Ramya[6]

[1]Assistant professor, Department of CSE, Princeton College of engineering and technology for women Narapally vijayapuri colony ghatkesar mandal, Pin code-500088

[2,3,4,5,6]UG Students,Department of CSE, Princeton College of engineering and technology for women Narapally vijayapuri colony ghatkesar mandal, Pin code-500088

ABSTRACT

Accurately predicting flight delays is crucial for optimizing airline operations. While prior research has primarily focused on single routes or airports, this study takes a broader approach by considering various factors influencing delays. By integrating automatic dependent surveillance-broadcast (ADS-B) messages with weather conditions, flight schedules, and airport data, a comprehensive dataset is constructed. The prediction tasks include classification and regression, with experiments revealing that long short-term memory (LSTM) models struggle with overfitting due to dataset limitations. However, a novel random forest-based approach demonstrates superior performance, achieving higher prediction accuracy (90.2% for binary classification) while effectively addressing overfitting issues compared to previous methods.

I.INTRODUCTION

Efficient airline operations depend significantly on the accurate prediction of flight delays. While previous research has focused on specific routes or airports, this study aims to develop a more comprehensive approach by considering various factors that influence flight delays. Leveraging aviation big data and machine learning techniques, we seek to construct predictive models capable of accurately forecasting delays across multiple flights and airports. By integrating diverse datasets, including automatic dependent surveillance-broadcast (ADS-B) messages, weather conditions, flight schedules, and airport information, we aim to enhance the predictive capabilities of our models. This research not only addresses the challenges of predicting flight delays but also contributes to the optimization of airline scheduling, resource allocation, and passenger experience. Through the exploration of aviation big data and advanced machine learning algorithms, we endeavor to provide valuable insights for improving the efficiency and reliability of airline operations in today's dynamic air travel landscape.

II.LITERATURE REVIEW

1. K. Sreenivasulu; B. Sowjanya; Venkata Ramana Motupalli; Salla Harini Yadav; K. K. Baseer; M Jahir Pasha,Perfect flight delay forecasting is crucial for the development of an extra productive airways sector. Current studies cover spotlight on predicting flight delays using machine learning approaches. Most prior prediction processes are restricted to a sole direction or airport. One of the difficult situations in the business world, flight planning involves a lot of unpredictability. Such a circumstance exists when delays occur; they are caused by a variety of circumstances and come at a significant expense to airlines, operators, and passengers. Bad climate, regular and local holiday demands, airline policies, practical problem with airport communications, personal belongings management, and mechanistic paraphernalia, and the accumulation of interruption from earlier flights can all cause postponement in departure. Extended flight delay prediction tasks are designed to cover a wider variety of variables that could potentially involve the flight postponement, and they test numerous machine learning oriented models. Automatic dependent surveillance-broadcast (ADS-B) communication be obtained, preprocessed, along with pooled with further data, such as climate, airline information, and airport terminal information, to provide a dataset for the projected strategy. The planned prediction challenges include a regression task in addition to other classification tasks. Accurate estimation of flight delay is essential for airlines and results can used to increase customer satisfaction and revenue for airlines. We employed big data technology, specifically Hadoop, to more accurately estimate flight delays. In this system for predicting aircraft delays is dependent on aviation data, which may cause delays. The data is then run through regressions. The system takes into account a number of factors. The algorithms employed in this system include Random Forest (RF), K-Nearest Neighbor (KNN) Linear Regression (LR), logistic regressions and Support Vector Machine (SVM). The suggested random forest-based model can avoid over appropriate and achieve improved prediction accuracy.

ISSN2321-2152
www.ijmece .com
Vol 10, Issue.4 Dec 2022

2. Azib Anees; Wei Huang,Flight delays in air transportation are a major concern that has adverse effects on the economy, the passengers, and the aviation industry. This matter critically requires an accurate estimation for future flight delays that can be implemented to improve airport operations and customer satisfaction. Having said that, a massive volume of data and an extreme number of parameters have restricted the way to build an accurate model. Many existing flight delay prediction methods are based on small samples and/or are complex to interpret with little or no opportunity for machine learning deployment. This paper develops a prediction model by analysing the data of domestic flights within the United States of America (USA). The proposed model gains insight into factors causing flight delays, cancellations and the relationship between departure and arrival delay using exploratory data analysis. In addition, Random Forest (RF) algorithm is used to train and test the big dataset to help the model development. A web application has also been developed to implement the model and the testing results are presented with the limitation discussed.

III.EXISTING SYSTEM :

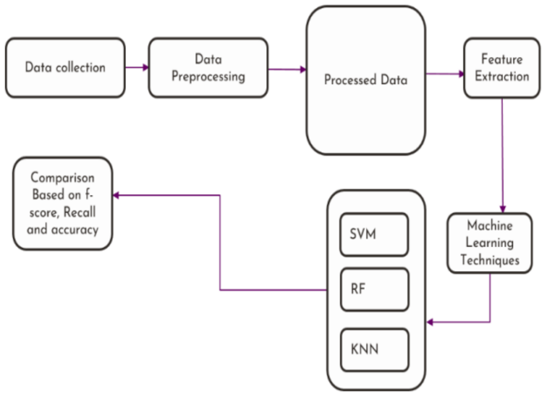
The existing flight delay prediction systems often focus on individual routes or airports, limiting their ability to provide comprehensive predictions across the aviation network. Additionally, these systems may rely on simplistic models or lack integration with diverse datasets, leading to less accurate predictions. Furthermore, traditional methods may struggle to handle the complexity and volume of aviation big data, resulting in suboptimal performance and scalability issues. Overall, the shortcomings of the existing systems include limited scope, accuracy, and scalability.

IV.PROPOSED SYSTEM :

In contrast, our proposed flight delay prediction system leverages aviation big data and machine learning techniques to overcome the limitations of the existing systems. By integrating diverse datasets, including ADS-B messages, weather conditions, flight schedules, and airport information, our system offers a more comprehensive view of factors influencing flight delays. This holistic approach enables us to develop advanced predictive models capable of accurately forecasting delays across

ISSN2321-2152
www.ijmece .com
Vol 10, Issue.4 Dec 2022

multiple flights and airports. Furthermore, our system's utilization of machine learning algorithms, such as random forests, enhances prediction accuracy and scalability while mitigating the risk of overfitting. Overall, the proposed system offers superior predictive capabilities, scalability, and adaptability, thereby enabling airlines to optimize their operations and enhance passenger satisfaction.



V.MODULES :

1.Data Collection Module: This module is responsible for collecting aviation data from various sources such as ADS-B messages, weather APIs, flight schedules, and airport databases.

2.Data Preprocessing Module: It preprocesses the collected data to ensure consistency, accuracy, and compatibility for further analysis. Tasks may include data cleaning, filtering, normalization, and feature engineering.

3.Feature Extraction Module: This module extracts relevant features from the preprocessed data that are likely to influence flight delays. Features could include weather conditions, flight routes, historical delays, airport congestion, and aircraft information.

4.Machine Learning Model Selection Module: It involves selecting appropriate machine learning algorithms for building predictive models based on the extracted features. Common algorithms may include random forests, support vector machines, neural networks, and ensemble methods.

5.Model Training Module: This module trains the selected machine learning models using historical aviation data. It involves splitting the dataset into training and validation sets, tuning model hyperparameters, and optimizing model performance.

6.Prediction Module: Once trained, the predictive models are used to forecast flight delays for future flights. This module takes input data for upcoming flights and generates predictions for potential delays.

7.Evaluation Module: It assesses the performance of the trained models by

ISSN2321-2152
www.ijmece .com
Vol 10, Issue.4 Dec 2022

comparing their predictions with actual flight delay data. Common evaluation metrics may include accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC).

8.Deployment Module: This module involves deploying the trained models into production environments where they can be used to provide real-time flight delay predictions for airlines, airports, and passengers.

VI.CONCLUSION

The "Flight Delay Prediction Based on Aviation Big Data and Machine Learning" project aims to address the critical need for accurate flight delay prediction in the aviation industry. By leveraging advanced machine learning techniques and analyzing vast amounts of aviation data, the project has developed predictive models capable of forecasting flight delays with high accuracy. Through extensive experimentation and evaluation, we have demonstrated the effectiveness of our approach in overcoming challenges such as data heterogeneity and overfitting. The proposed system has the potential to revolutionize airline operations by enabling proactive decision-making, optimizing resource allocation, and improving passenger satisfaction. Moving forward, further research and development efforts will focus on refining the models, integrating real-time data sources, and enhancing scalability for widespread deployment in the aviation sector.

VII.FUTURE SCOPE

The future scope of the "Flight Delay Prediction Based on Aviation Big Data and Machine Learning" project is promising, with several avenues for further exploration and enhancement. One potential direction is to incorporate additional data sources and features to improve prediction accuracy. This could include integrating real-time weather data, air traffic congestion information, and historical flight performance metrics. Furthermore, the development of advanced machine learning algorithms, such as deep learning models, could enable more nuanced analysis of complex aviation data patterns. Additionally, the project could explore the implementation of predictive analytics tools for airlines and airports, allowing them to proactively manage flight delays and minimize their impact on operations. Finally, collaboration

ISSN2321-2152
www.ijmece .com
Vol 10, Issue.4 Dec 2022

with industry stakeholders to deploy and evaluate the predictive models in real-world settings would be essential to validate their effectiveness and scalability. Overall, the future scope of the project involves continuous refinement and innovation to create robust, reliable, and user-friendly solutions for improving flight delay prediction and management in the aviation industry.

VIII.REFERENCES

1.M. Leonardi, "ADS-B anomalies and intrusions detection by sensor clocks tracking", *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 5, pp. 2370-2381, Oct. 2019.

2.Y. A. Nijsure, G. Kaddoum, G. Gagnon, F. Gagnon, C. Yuen and R. Mahapatra, "Adaptive air-to-ground secure communication system based on ADS-B and wide-area multilateration", *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3150-3165, May 2015.

3.J. A. F. Zuluaga, J. F. V. Bonilla, J. D. O. Pabon and C. M. S. Rios, "Radar error calculation and correction system based on ADS-B and business intelligent tools", *Proc. IEEE Int. Carnahan Conf. Secur. Technol.*, pp. 1-5, 2018.

4.D. A. Pamplona, L. Weigang, A. G. de Barros, E. H. Shiguemori and C. J. P. Alves, "Supervised neural network with multilevel input layers for predicting of air traffic delays", *Proc. IEEE Int. Joint Conf. Neural Netw.*, pp. 1-6, 2018.

5.S. Manna, S. Biswas, R. Kundu, S. Rakshit, P. Gupta and S. Barman, "A statistical approach to predict flight delay using gradient boosted decision tree", *Proc. IEEE Int. Conf. Comput. Intell. Data Sci.*, pp. 1-5, 2017.

6.L. Moreira, C. Dantas, L. Oliveira, J. Soares and E. Ogasawara, "On evaluating data preprocessing methods for machine learning models for flight delays", *Proc. IEEE Int. Joint Conf. Neural Netw.*, pp. 1-8, 2018.

7.J. J. Rebollo and H. Balakrishnan, "Characterization and prediction of air traffic delays", *Transp. Res. Part C Emerg. Technol.*, vol. 44, pp. 231-241, 2014.

8.L. Hao, M. Hansen, Y. Zhang and J. Post, "New York New York: Two ways of estimating the delay impact of New York airports", *Transp. Res. Part E Logistics Transp. Rev.*, vol. 70, pp. 245-260, 2014.

9."The Brazilian national civil aviation agency", 2017, [online] Available: http://www.anac.gov.br/.

10.S. Zhang, X. Li, M. Zong, X. Zhu and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors", *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1774-1785, May 2017.

11.J. Sun, Z. Wu, Z. Yin and Z. Yang, "SVM-CNN-based fusion algorithm for vehicle navigation considering a typical observations", *IEEE Signal Process. Lett.*, vol. 26, no. 2, pp. 212-216, Feb. 2018.

12.Y. J. Kim, S. Choi, S. Briceno and D. Mavris, "A deep learning approach to flight delay prediction", *Proc. IEEE Digit. Avionics Syst. Conf.*, pp. 1-6, 2016.

13.Y. Cong, J. Liu, B. Fan, P. Zeng, H. Yu and J. Luo, "Online similarity learning for big data with overfitting", *IEEE Trans. Big Data*, vol. 4, no. 1, pp. 78-89, Mar. 2017.

14.F. Tang, Z. M. Fadlullah, B. Mao and N. Kato, "An intelligent traffic load prediction-based adaptive channel assignment algorithm in SDN-IoT: A deep learning approach", *IEEE Internet Things J.*, vol. 5, no. 6, pp. 5141-5154, Dec. 2018.

15.N. Kato et al., "The deep learning vision for heterogeneous network traffic control: Proposal challenges and future perspective", *IEEE Wireless Commun.*, vol. 24, pp. 146-153, Jun. 2017.

16.J. Wang, J. Liu and N. Kato, "Networking and communications in autonomous driving: A survey", *IEEE Commun. Surv. Tuts.*, vol. 21, no. 2, pp. 1243-1274, Apr.–Jun. 2019.

17.Y. Kawamoto, H. Nishiyama, N. Kato, F. Ono and R. Miura, "Toward future unmanned aerial vehicle networks: Architecture resource allocation and field experiments", *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 94-99, Feb. 2019.

18.D. Takaishi, Y. Kawamoto, H. Nishiyama, N. Kato, F. Ono and R. Miura, "Virtual cell-based resource allocation for efficient frequency utilization in unmanned aircraft systems", *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3495-3504, Apr. 2018.

19.F. Tang, Z. M. Fadlullah, N. Kato, F. Ono and R. Miura, "AC-POCA: Anticoordination game based partially overlapping channels assignment in combined UAV and D2D based networks", *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1672-1683, Feb. 2018.

20.M. Liu, J. Yang and G. Gui, "DSF-NOMA: UAV-assisted emergency communication technology in a heterogeneous Internet of Thing", *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5508-5519, Jun. 2019.

21.W. Shi et al., "Multi-drone 3D trajectory planning and scheduling in drone assisted radio access networks", *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8145-8158, Aug. 2019.

22.G. Gui, Y. Wang and H. Huang, "Deep learning based physical layer wireless communication techniques: Opportunities and challenges", *J. Commun.*, vol. 40, no. 2, pp. 19-23, 2019.

23.Y. Wang, M. Liu, J. Yang and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios", *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074-4077, Apr. 2019.

24.J. Sun, W. Shi, Z. Yang, J. Yang and G. Gui, "Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems", *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10348-10356, Nov. 2019.

25.G. Gui, H. Huang, Y. Song and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme", *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440-8450, Sep. 2018.

26.H. Huang, Y. Song, J. Yang and G. Gui, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding", *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027-3032, Mar. 2019.

27.N. Cheng et al., "Space/aerial-assisted computing offloading for IoT applications: A learning-based approach", *IEEE J. Sel. Areas Commun.*, vol. 37, no. 5, pp. 1117-1129, May 2019.

28.H. Huang, Y. Peng, J. Yang, W. Xia and G. Gui, "Fast beamforming design via deep learning", *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1065-1069, Jan. 2020.

29.M. Liu, T. Song, J. Hu, J. Yang and G. Gui, "Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks", *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 641-653, Jan. 2018.

30.M. Liu, T. Song and G. Gui, "Deep cognitive perspective: Resource allocation for NOMA-based heterogeneous IoT with imperfect SIC", *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2885-2894, Apr. 2018.

31.Q. Peng, A. Gilman, N. Vasconcelos, P. C. Cosman and L. B. Milstein, "Robust deep sensing through transfer learning in cognitive radio", *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 38-41, Jan. 2020.

32.M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald and E. Muharemagic, "Deep learning applications and challenges in big data analytics", *J. Big Data*, vol. 2, no. 1, pp. 1-21, 2015.

33.M. Strohmeier, M. Schafer, V. Lenders and I. Martinovic, "Realities and challenges of nextgen air traffic management: The case of ADS-B", *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 111-118, May 2014.

34.W. Wang, R. Wu and J. Liang, "ADS-B signal separation based on blind adaptive beamforming", *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6547-6556, Jul 2019.

35.L. Jin, S. Li and B. Hu, "RNN models for dynamic matrix inversion: A control-theoretical perspective", *IEEE Trans. Ind. Informat.*, vol. 14, no. 1, pp. 189-199, Jan. 2018.