Because currently I have not read enough scientific literature in the field, detailed methodology and objectives of the research might not be settled, and the introduction can be seen as a rough draft with basic frameworks of the outcome and approaches of the research. Also, since the literature review does not have much content and most a second-hand conclusion from other literature reviews, I integrate it into the introduction.

Topic: TBD

1. Introduction

Air transportation is an essential component in a national transportation system, while delay is one of the most significant performance indicators of aviation service. With the conclusion of COVID-19 pandemic, there is a significant recovering trend in demands of flights worldwide. Airlines across mainland China completed 4.611 million flights in 2019, marking a 6.1% increase in last year’s basis. And it can be foreseen that the number will recover quickly after the end of pandemic [1]. Leisure seats ordering in Europe and North America are also back to post-pandemic level [2]. However, flight delay seems to be a long-lasting issue in air-transportation systems. In 2013, Europe saw 36% of its flights delayed by over five minutes, while in the United States, 31.1% of flights experienced delays exceeding 15 minutes. Additionally, in Brazil, 16.3% of flights were either canceled or faced delays of more than 30 minutes [3, 5]. Meanwhile, flight delays can have multiple influences, including but not limited to economics, environment, airlines, airports. Study in 2013 suggests that $17.6 million increase in the U.S net worth can be made by merely 10% of flight delay, while with 30% of flight delay, the value would be $38.5 million [4]. Study in 2018 [5] indicates that in 2017, extra emission due to flight delay reached an estimation of 5529 tons, together with roughly 1,752,937 L more fuel consumed. Given the uncertainty of occurrence, passengers will have to spend extra hours on traveling, which leads to higher cost [6, 7]. On the other hand, as result of flight delays, airlines suffer from penalties and excess operational spendings for retention in airport, crew cost [8, 9, 10]. It is obvious that flight delays will have negative impact, on the other hand, successfully predicting the delay of flights can boost decision making and reduce loss via advance actions. All above, the topic of flight delays has high relevance to the social community.

Different types of efforts have been made in the field. In most cases, delay can be described as ‘the period by which a flight is late or postponed’ [11], which makes it deviation of real arrival or departure time from scheduled time [12]. In a literature review [11], it is concluded that problems or information of interests in the field are in three categories, which are delay propagation, New (root) delay, and cancellation. In root delay and cancellation, several topics are covered, including analyzing condition for occurrence, distribution, probabilistic models to predict occurrence time and location and estimate durations. As for delay propagation, impact of flight delay for other flights in the same airline, airport, or even other airports are studied based on assumption that a delay has already happened [11]. As for methods, different methods are applied according to the problems and scope, while there is a significant increase in using machine learning methods in the late 2000s [11], which indicates trending in concentration on machine learning utilization.

Specifically in machine learning research in the field, different data preprocessing techniques and machine learning algorithms together with performance metrics are utilized. It is spotted that most studies are based on small sample data, which indicates tailored configuration and specific compatibility [13], while most of studies also focus on classification problem [13]. In data preprocessing, feature selections notably progress in recent studies [13]. When it comes to algorithms, choices include but are not limited to k-Nearest Neighbor, neural networks, SVM, fuzzy logic, and random forests [11], convolutional neural networks, decision trees or ensemble techniques [13]. Among these, random forests and gradient boosting machines are found effective [13].

Using existing studies as reference and fundament, this paper will test four algorithms, Logistic Regression, SVM, ANN, GP, to predict whether delay or not based on conditions. Data source comes from open-source Kaggle dataset, which is inspired by the dataset from U.S. Department of Transportation. Feature engineering, data separation into training, testing, validation will be applied in data preprocessing to enhance efficiency and reduce overfitting potential. Performance metrics including confusion matrix, ROC curve, AUC value will be calculated to measure model performance. Since utilization of Genetic Computing in this field has not been researched, this paper may contribute by improving algorithms and other techniques in flight delay prediction.

Reference

[1] Z. Wang *et al.*, “Distribution prediction of strategic flight delays via machine learning methods,” *Sustainability*, vol. 14, no. 22, p. 15180, Nov. 2022. doi:10.3390/su142215180

[2] T. Stalnaker, K. Usman, G. Alport, A. Buchanan, and A. Taylor, “Post-covid trends in Airline Economic Analysis 2022–2023,” Oliver Wyman - Impact-Driven Strategy Advisors, https://www.oliverwyman.com/our-expertise/insights/2023/may/airline-economic-analysis-2022-2023.html (accessed Mar. 21, 2024).

[3] EUROCONTROL. CODA Digest - Delays to Air Transport in Europe. Technical report, https://www.eurocontrol.int/articles/coda-publications, 2017.

[4] Peterson, E.B.; Neels, K.; Barczi, N.; Graham, T. The economic cost of airline flight delay. J. Transp. Econ. Policy 2013, 47, 107–121.

[5] Dissanayaka, D.; Adikariwattage, V.; Pasindu, H. Evaluation of Emissions from Delayed Departure Flights at Bandaranaike International Airport (BIA). In Proceedings of the 11th Asia Pacific Transportation and the Environment Conference (APTE 2018), Malang, Indonesia, 18–19 October 2018 ; Atlantis Press: Paris, France, 2019; pp. 143–146. [CrossRef]

[6] P. Balakrishna, R. Ganesan, and L. Sherry. Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: A case-study of Tampa Bay departures. Transportation Research Part C: Emerging Technologies, 18(6):950–962, Dec. 2010. ISSN 0968-090X.

[7] P. Fleurquin, B. Campanelli, V. Eguiluz, and J. Ramasco. Trees of reactionary delay: Addressing the dynamical robustness of the US air transportation network. In SIDs 2014 - Proceedings of the SESAR Innovation Days, 2014.

[8] R. Britto, M. Dresner, and A. Voltes. The impact of flight delays on passenger demand and societal welfare. Transportation Research Part E: Logistics and Transportation Review, 48(2):460–469, Mar. 2012. ISSN 13665545.

[9] Y. Tu, M. O. Ball, and W. S. Jank. Estimating flight departure delay distributions—a statistical approach with long-term trend and short-term pattern. Journal of the American Statistical Association, 103(481):112–125, 2008.

[10] J. Evans, S. Allan, and M. Robinson. Quantifying delay reduction benefits for aviation convective weather decision support systems. In Conference on Aviation, Range, and Aerospace Meteorology, pages 39–70, 2004.

[11] L. Carvalho *et al.*, “On the relevance of data science for flight delay research: a systematic review,” *Transport Reviews*, vol. 41, no. 4, pp. 1–30, Dec. 2020, doi: <https://doi.org/10.1080/01441647.2020.1861123>.

[12] F. Wieland. Limits to growth: results from the detailed policy assessment tool [air traffic congestion]. In 16th DASC. AIAA/IEEE Digital Avionics Systems Conference. Reflections to the Future. Proceedings, volume 2, pages 9.2–1–9.2–8 vol.2, Oct. 1997.

[13] K. Kiliç and J. M. Sallan, “Study of Delay Prediction in the US Airport Network,” *Aerospace*, vol. 10, no. 4, p. 342, Apr. 2023, doi: https://doi.org/10.3390/aerospace10040342.

‌ [1]

E. Mueller and G. Chatterji, “Analysis of Aircraft Arrival and Departure Delay Characteristics,” *AIAA’s Aircraft Technology, Integration, and Operations (ATIO) 2002 Technical Forum*, Oct. 2002, doi: https://doi.org/10.2514/6.2002-5866.

‌