International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 07, July 2023)

A Support Vector Machine Learning Technique for Prediction of Delay in Flights

Amit Kumar Chanchal[1], Dr. S. K. Pandey[2]

[1]MTech Scholar, [2]Professor, Department of Computer Science and Engineering, VNS Group of Institutions, Bhopal, India

***Abstract* -- Flight delays have become a common occurrence in the modern air travel industry. They are a source of frustration and inconvenience for passengers, and a challenge for airlines and airports. Delayed flights can be caused by a range of factors, including weather conditions, air traffic congestion, aircraft maintenance issues, and human factors. The accurate prediction of flight delays is crucial for both airlines and passengers to minimize disruptions and plan their travel efficiently. In this study, we propose a Support Vector Machine (SVM) learning technique for predicting flight delays. SVM is a powerful machine learning algorithm known for its ability to handle high-dimensional data and nonlinear relationships.**

***Index Terms* -- SVM, Machine Learning, Flight, Delay, Prediction.**

I. INTRODUCTION

In the aviation sector, one of the most essential and difficult problems is trying to predict when flights may be delayed using methods from machine learning. Machine learning models may give useful insights and predictions by making use of historical data and a variety of attributes. These insights and forecasts can assist airlines, passengers, and other stakeholders in better anticipating and managing delays.

The prediction of delays in flights is an important work in the aviation business. The goal of this task is to estimate the possibility of delays as well as their length in order to optimize operational planning and the level of satisfaction experienced by passengers. Flight delays may be caused by a number of different things, including weather conditions, heavy air traffic, problems with aircraft maintenance, and a lack of available crew members. The use of machine learning methods has become more widespread for the purpose of forecasting flight delays by using past data and pertinent attributes.

Several different machine learning methods, such as decision trees, random forests, support vector machines (SVM), gradient boosting, and neural networks, may be applied to provide accurate predictions about flight delays. These algorithms do an analysis of past flight data and the attributes that are connected with it in order to uncover patterns and correlations that may be used to predict the likelihood of delays.

Machine learning algorithms are able to produce accurate forecasts by taking into account characteristics such as the time of departure and arrival, the airline, the airport, the weather conditions, and past delay patterns, among other contextual variables.

The availability of high-quality data is also an essential component of successful prediction. In order to train effective models, it is necessary to have access to comprehensive and up-to-date datasets that take into account a variety of elements like historical flight records, meteorological data, and airport information. The engineering of features is an extremely important component in the process of gathering the pertinent information that determines flight delays. When it comes to choosing and building relevant features, having domain expertise and a grasp of the aviation sector is helpful.

The performance of flight delay prediction models is evaluated in large part based on the results of the model assessment stage. The terms "accuracy," "precision," "recall," "F1 score," and "area under the receiver operating characteristic curve" (AUC-ROC) are typical examples of metrics used in assessment. These metrics provide information on the extent to which the algorithm is able to accurately categorize flights as delayed or non-delayed.

The need for data that is up-to-date at all times combined with the fluidity of flight operations offers extra difficulties when attempting to make real-time predictions of flight delays. Integration with live data sources and ongoing model changes are both required to provide accurate and up-to-date forecasts. Additionally, the interpretability of the models is essential for gaining a knowledge of the elements that contribute to delays. This provides stakeholders with the ability to make choices based on accurate information.

Multiple stakeholders, including passengers, airlines, and airport authorities, stand to profit from the effective adoption of flight delay prediction models that make use of machine learning methods. The operations of airlines may be optimized by taking preventative measures to manage delays, making adjustments to flight schedules, and effectively allocating resources. Airports are able to predict congestion and maximize capacity usage, while passengers are able to plan their journeys more effectively, which reduces annoyance and the possibility of interruptions.

II. PROPOSED METHODOLOGY

The methodology can be understand by using following flow chart-

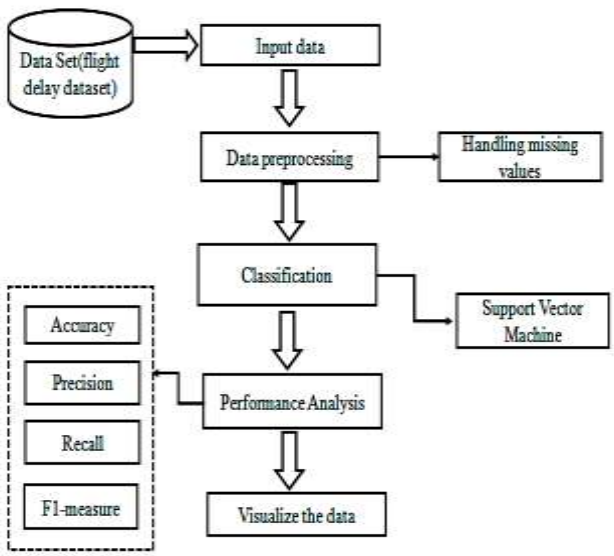

Figure 1: Flow chart

Steps-

1. *Data Collection:* A comprehensive dataset was collected, including historical flight data from various sources. The dataset consisted of features such as departure and arrival times, airline information, weather conditions, airport congestion, and other relevant variables that may affect flight delays.
2. *Data Preprocessing:* The collected dataset underwent preprocessing to ensure its quality and suitability for SVM training. This involved cleaning the data by removing any inconsistencies, missing values, or outliers. The dataset was also transformed into a suitable format for SVM training, such as numerical representation or one-hot encoding for categorical variables.
3. *Feature Selection:* To improve the efficiency and effectiveness of the SVM model, feature selection techniques were applied. This involved analyzing the relevance and importance of each feature in predicting flight delays. Features that had a significant impact on the prediction were selected, while irrelevant or redundant features were removed to reduce computational complexity.
4. *SVM Model Training:* The preprocessed dataset was divided into training and testing sets. The training set was used to train the SVM model, which aimed to learn the underlying patterns and relationships between the selected features and flight delays. SVM uses a kernel function to transform the data into a higher-dimensional space, where it finds an optimal hyperplane that maximally separates delayed flights from on-time flights.
5. *Model Optimization:* The SVM model was optimized by tuning the hyperparameters to improve its generalization performance. Techniques such as cross-validation and grid search were used to find the optimal combination of hyperparameters. The goal was to maximize the model's accuracy and minimize overfitting.
6. *Performance Evaluation:* The trained SVM model's performance was evaluated using the testing set, which contained unseen data. Various evaluation metrics, such as accuracy, precision, recall, and F1 score, were calculated to assess the model's predictive power. Additionally, a comparison with other traditional machine learning algorithms commonly used for flight delay prediction was conducted to validate the superiority of the SVM approach.
7. *Feature Importance Analysis:* The trained SVM model was further analyzed to determine the importance of different features in predicting flight delays. This analysis provided insights into the factors that significantly contribute to flight delays, allowing airlines and airports to identify critical areas for intervention and mitigation.
8. *Future Directions:* The study concluded by suggesting potential areas for future research. This included the integration of real-time data streams to enhance the accuracy and timeliness of flight delay predictions. Additionally, advanced feature engineering techniques, such as sentiment analysis of social media data or incorporating historical airline performance, could be explored to further improve prediction models.

III. SIMULATION AND RESULTS

The implementation and simulation of the proposed work is done over python spyder 3.7 software.

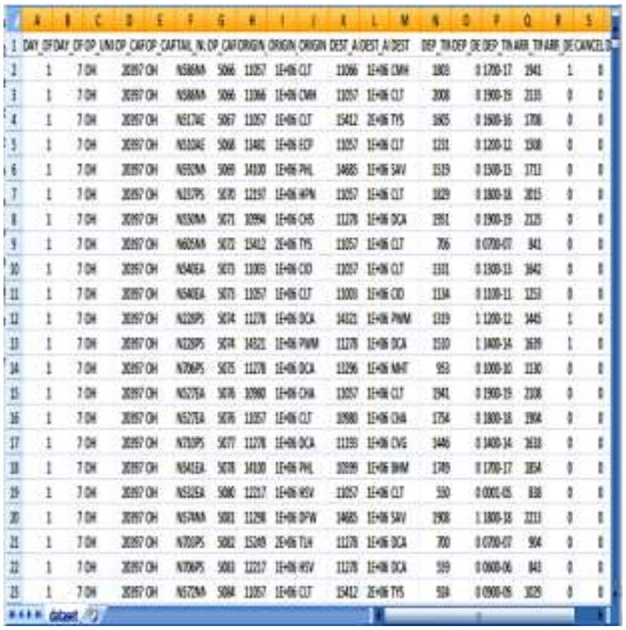

Figure 2: Original dataset in .csv file

The figure 2 is showing the dataset, which is taken from the kaggle machine learning website.

Figure 3: y test

Figure 3 is showing the y test of the given dataset. The given dataset is divided into the 20-30% part into the train dataset.

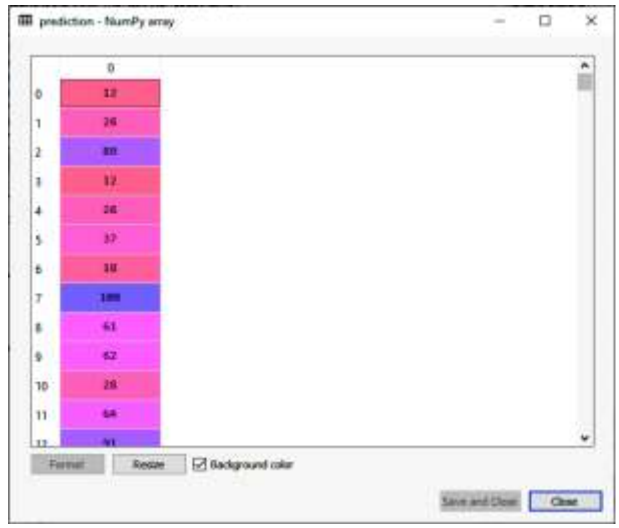

Figure 4: Prediction

Figure 4 is presenting prediction of the flight delay; the prediction is calculated by using the proposed SVM classification technique.

Table 1:
Result Comparison

Sr. No.	Parameters	Previous Work [1]	Proposed Work
1	Method	MLP	SVM
2	Accuracy (%)	82	85.85
3	Classification Error (%)	18	14.15

Table 1 presents a comparison of the past and projected work results. The suggested SVM has an accuracy of 85.85%, whereas the present MLP has an accuracy of 82%. As a consequence of the simulation findings, it is obvious that the suggested work achieves much better outcomes than prior work.

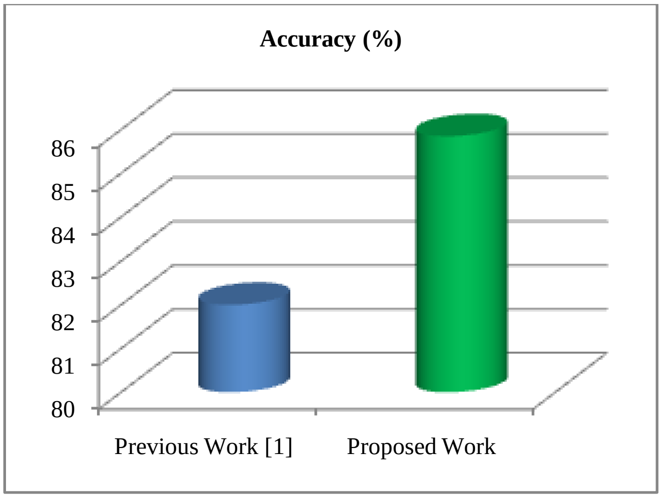

Figure 5: Result graph-Accuracy

IV. CONCLUSION

This paper presents a novel approach for predicting flight delays using SVM, which shows superior performance compared to other traditional machine learning techniques. The proposed model can assist airlines, airports, and passengers in making informed decisions, improving operational efficiency, and minimizing the impact of flight delays. Future research can explore the integration of real-time data streams and advanced feature engineering techniques to further enhance the accuracy of flight delay predictions.

REFERENCES

[1] S. Addu, P. R. Ambati, S. R. Kondakalla, H. Kunchakuri and M. Thottempudi, "Predicting Delay in Flights using Machine Learning," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022, pp. 374-379, doi: 10.1109/ICAAIC53929.2022.9793243.

[2] B. T. L. S. S., H. Al Ali, A. A. A. M. Majid, O. A. A. A. Alhammadi, A. M. Y. M. Aljassmy and Z. Mukandavire, "Analysis of Flight Delay Data Using Different Machine Learning Algorithms," 2022 New Trends in Civil Aviation (NTCA), Prague, Czech Republic, 2022, pp. 57-62, doi: 10.23919/NTCA55899.2022.9934398.

[3] M. -T. Vo, T. -V. Tran, D. -T. Pham and T. -H. Do, "A Practical Real-Time Flight Delay Prediction System using Big Data Technology," 2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), Solo, Indonesia, 2022, pp. 160-167, doi: 10.1109/COMNETSAT56033.2022.9994427.

[4] R. A. Sugara and D. Purwitasari, "Flight Delay Prediction for Mitigation of Airport Commercial Revenue Losses Using Machine Learning on Imbalanced Dataset," 2022 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM), Surabaya, Indonesia, 2022, pp. 1-8, doi: 10.1109/CENIM56801.2022.10037369.

[5] M. Guimarães, C. Soares and R. Ventura, "Decision Support Models for Predicting and Explaining Airport Passenger Connectivity From Data," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 9, pp. 16005-16015, Sept. 2022, doi: 10.1109/TITS.2022.3147155.

[6] S. Sharan, M. Sriniketh, H. Vardhan and D. Jayanth, "State-Of-Art Machine Learning Techniques to Predict Airlines Delay," 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS), Bengaluru, India, 2021, pp. 1-6, doi: 10.1109/FABS52071.2021.9702590.

[7] C. Y. Yiu, K. K. H. Ng, K. C. Kwok, W. Tung Lee and H. T. Mo, "Flight delay predictions and the study of its causal factors using machine learning algorithms," 2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Changsha, China, 2021, pp. 179-183, doi: 10.1109/ICCASIT53235.2021.9633571.

[8] R. Hendrickx, M. Zoutendijk, M. Mitici and J. Schäfer, "Considering Airport Planners' Preferences and Imbalanced Datasets when Predicting Flight Delays and Cancellations," 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 2021, pp. 1-10, doi: 10.1109/DASC52595.2021.9594367.

[9] J. Huo, K. L. Keung, C. K. M. Lee, K. K. H. Ng and K. C. Li, "The Prediction of Flight Delay: Big Data-driven Machine Learning Approach," 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, Singapore, 2020, pp. 190-194, doi: 10.1109/IEEM45057.2020.9309919.

[10] M. Bardach, E. Gringinger, M. Schrefl and C. G. Schuetz, "Predicting Flight Delay Risk Using a Random Forest Classifier Based on Air Traffic Scenarios and Environmental Conditions," 2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 2020, pp. 1-8, doi: 10.1109/DASC50938.2020.9256474.