International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

FLIGHT DELAY ANALYSIS AND PREDICTION

[1] B Jyothi, [2] P Meghana, [3] C Shravani, [4] K Kamaleshwar Reddy
[1]AssistantProfessor, [234]Students
Department of Data Science
Siddhartha Institute of Technology & Sciences, Narapally
Bjyothi.cse@siddhartha.org.in, 23TQ1A6712@siddhartha.co.in, 23TQ1A6737@siddhartha.co.in, 23TQ1A6747@siddhartha.co.in,

ABSTRACT

Flight delays are a common problem in the aviation industry and can affect passengers, airlines, airport operations, and overall transportation efficiency. Delays may occur due to several factors, including weather conditions, air traffic congestion, aircraft-related issues, airport operations, and scheduling constraints. Analyzing historical flight data can help identify patterns associated with delays and support better operational planning.

The Flight Delay Analysis and Prediction system is designed to analyze historical flight records and identify the factors that contribute to flight delays. The system uses information such as airline, origin airport, destination airport, scheduled departure time, arrival time, flight duration, weather conditions, and previous delay information. Data preprocessing techniques are applied to clean and prepare the dataset for analysis. The system performs sexploratory data analysis to identify delay patterns across airlines, airports, routes, time periods, and other relevant factors. Visualization techniques such as bar charts, line graphs, heatmaps, and dashboards can be used to represent delay distributions and trends. These visualizations help users understand which factors are associated with higher or lower delay frequencies.

A Machine Learning model is used to predict whether a particular flight is likely to experience a delay based on available input features. Classification algorithms can be trained using historical flight records and evaluated using suitable performance metrics. The system can display the prediction along with relevant analytical information to help users understand the result.

Overall, the proposed system combines historical flight-data analysis with predictive modeling to provide useful insights into flight delays. It can support airlines, airport analysts, researchers, and other authorized users in understanding delay patterns and improving operational planning. Future enhancements can include real-time flight data, live weather information, advanced time-series models, route-specific prediction, and real-time delay alerts.

I. INTRODUCTION

Air transportation is an important part of modern transportation systems, connecting cities and countries efficiently. Millions of passengers travel by air every day, making timely flight operations important for both passengers and airlines. However, flight delays can disrupt travel schedules and create operational challenges for airports and airlines.

Flight delays can be caused by many factors, including unfavorable weather, air traffic congestion, aircraft maintenance, late arrival of previous flights, airport capacity, crew availability, and operational issues. Some factors are predictable to a certain extent, while others can change rapidly. Understanding historical delay patterns can therefore provide useful information for planning and analysis.

Traditional flight monitoring systems primarily focus on displaying flight schedules, statuses, and operational information. Although these systems are useful for tracking individual flights, they may not provide detailed analytical information about historical delay patterns. Analyzing large volumes of historical flight records manually can also be time-consuming.

Data analytics provides methods for examining flight information and identifying relationships between different variables and delays. Analysts can compare delay rates across airlines, airports, routes, months, days, and time periods. Visualization makes these relationships easier to understand and allows users to identify recurring patterns.

Machine Learning provides an additional approach by learning patterns from historical flight data and using them to predict potential delays. A prediction model can use features such as airline, route, departure time, weather, and previous operational information. The proposed Flight Delay Analysis and Prediction system combines data preprocessing, exploratory analysis, visualization, and Machine Learning to provide an integrated platform for understanding and predicting flight delays.

II. LITERATURE SURVEY

Flight delay analysis has been studied extensively because delays affect airline efficiency, airport operations, passenger satisfaction, and transportation planning. Historical flight datasets generally contain information about scheduled and actual departure times, arrival times, airlines, airports, routes, and delay durations. Researchers use such datasets to identify the causes and patterns of flight delays.

Statistical analysis has traditionally been used to understand relationships between flight characteristics and delays. Researchers can calculate average delay durations, delay frequencies, and cancellation rates across airlines, airports, routes, and time periods. Such descriptive statistics provide an overview of operational performance and help identify areas requiring further analysis.

Weather conditions are frequently considered an important factor in flight operations. Rain, storms, fog, strong winds, and other adverse weather conditions can affect airport capacity and flight schedules. Integrating weather-related variables with flight records can provide additional information for analyzing delay patterns.

Data preprocessing is an important stage in flight-delay analysis because real-world datasets may contain missing values, inconsistent timestamps, duplicate records, and categorical variables. Cleaning the data and transforming features into suitable formats helps improve the quality of analytical results. Feature engineering can also create useful attributes such as departure hour, day of week, month, route, and previous-flight delay.

Machine Learning techniques have been widely explored for flight-delay prediction. Classification algorithms can predict whether a flight may be delayed, while regression models can estimate the expected delay duration. Algorithms such as Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, and other machine-learning approaches can be evaluated using historical data and suitable performance metrics.

Recent systems combine data analytics, visualization, and predictive modeling to provide more comprehensive flight-delay decision support. Interactive dashboards can display delay trends while prediction models provide estimates for future flights. The proposed Flight Delay Analysis and Prediction system follows this approach by combining historical flight-data analysis, delay-factor analysis, visualization, and Machine Learning-based prediction.

III. SYSTEM ANALYSIS

The Flight Delay Analysis and Prediction system is designed to process historical flight data and generate analytical and predictive insights about flight delays. The system accepts information such as airline, origin airport, destination airport, scheduled departure time, arrival time, flight duration, weather conditions, and delay status. The input data is first validated to ensure that required fields and supported formats are available. A preprocessing module handles missing values, duplicate records, inconsistent timestamps, and categorical data. Feature engineering is performed to derive useful attributes such as departure hour, day of week, month, route, and flight duration. The analytics engine calculates metrics such as total flights, delayed flights, average delay duration, and delay percentage. Comparative analysis is performed across airlines, airports, routes, and time periods. Visualization components present delay patterns through charts, graphs, heatmaps, tables, and KPI cards. A Machine Learning model is trained using historical flight records and selected features. The model predicts whether a new flight is likely to experience a delay based on its available attributes. Model performance is evaluated using suitable classification or regression metrics. The system presents prediction results and historical analytical information through an interactive dashboard.

Existing System

Existing flight information systems mainly focus on providing flight schedules, status updates, departure information, arrival information, and basic delay notifications. These systems are useful for passengers and operational teams who need information about individual flights. However, basic flight-status systems may not provide detailed historical analysis of delay patterns. Analysts may need to use spreadsheets or separate databases to examine delay frequencies and average delay durations. Comparing airlines, airports, routes, and different time periods can require additional manual processing. Historical datasets may also contain missing or inconsistent information that requires preprocessing before analysis. Traditional systems generally focus on monitoring current operations rather than predicting future delays. Some analytical platforms may provide reports but have limited predictive capabilities. Users may also need separate tools for visualization and Machine Learning analysis. Identifying relationships between weather, departure time, airport congestion, and delays can be difficult without an integrated analytical platform. Manual analysis of large flight datasets can consume significant time and resources. Therefore, an integrated system combining flight-data analysis, visualization, and delay prediction can provide additional analytical capabilities.

Disadvantages of Existing System

- Limited historical flight-delay analysis.
- Focus mainly on current flight status and schedules.
- Manual analysis can be time-consuming.
- Difficult to analyze large flight datasets manually.
- Limited prediction of future flight delays.
- Separate tools may be required for advanced analytics.
- Difficult to compare multiple airlines and airports.
- Limited visualization of delay patterns.
- Weather and operational factors may not be analyzed together.
- Data preprocessing may require manual effort.
- Limited route-specific analytical insights.
- Static reports may not provide interactive analysis.

Proposed System

The proposed Flight Delay Analysis and Prediction system provides an integrated platform for analyzing historical flight data and predicting potential delays. Users can upload historical flight datasets containing information about airlines, airports, routes, schedules, weather conditions, and delay status. The system validates the dataset and performs preprocessing to handle missing values, duplicates, inconsistent timestamps, and categorical attributes. Feature engineering creates useful variables such as departure hour, day of week, month, route, and previous delay information when available. The analytics module calculates key metrics including total flights, delayed flights, delay percentage, and average delay duration. Users can compare delay patterns across airlines, airports, routes, and different time periods. Interactive charts and dashboards present these analytical results in an understandable format. A Machine Learning module trains a predictive model using historical flight information and selected features. When new flight details are provided, the model generates an estimated probability or classification of potential delay according to its trained output. The system can display the prediction along with relevant flight and analytical information. Model performance is evaluated using appropriate metrics before deployment. The proposed system therefore combines data analysis, visualization, and predictive modeling in a single platform.

Advantages of Proposed System

- Provides centralized flight-delay analysis.
- Automates data preprocessing and analysis.
- Identifies delay patterns across airlines.
- Supports airport and route comparison.
- Provides historical delay trends.
- Uses Machine Learning for delay prediction.
- Provides interactive dashboards and visualizations.
- Helps analyze time-based delay patterns.
- Can incorporate weather-related features.
- Reduces manual data-analysis effort.
- Supports analytical reporting.
- Provides a foundation for real-time prediction enhancements.

IV. METHODOLOGY

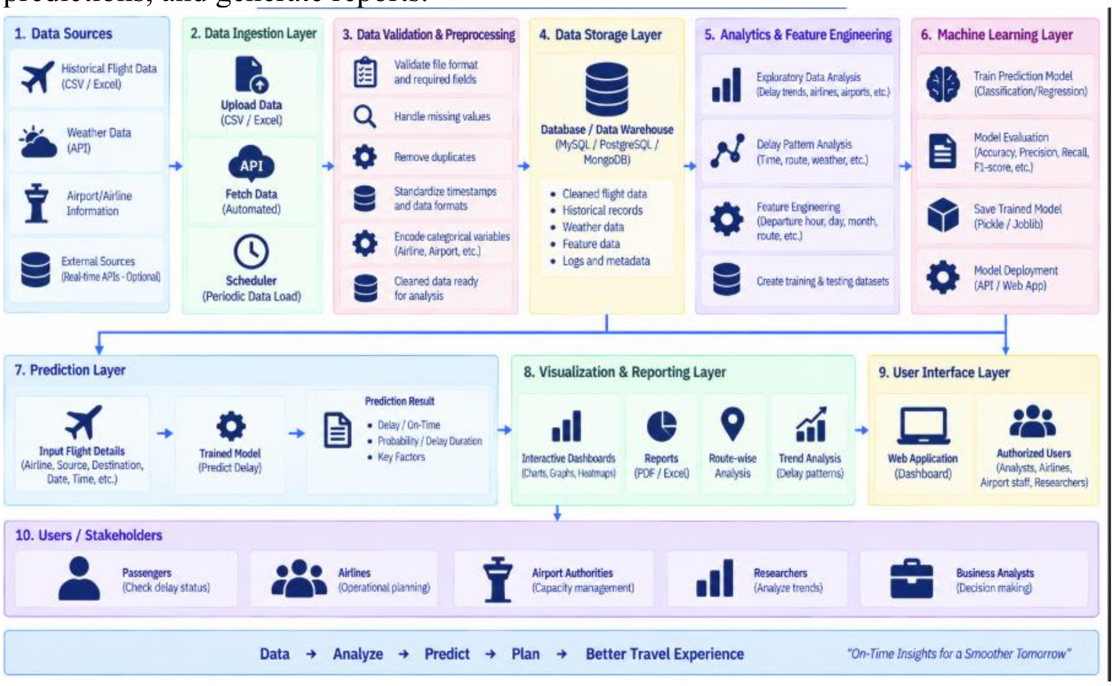
The methodology begins by collecting historical flight records from an authorized dataset or data source. The collected data is validated to check file format, required fields, data types, and overall completeness. Data preprocessing is performed to remove duplicate records and handle missing, invalid, or inconsistent values. Flight timestamps are standardized and categorical variables such as airline and airport are transformed into suitable formats. Feature engineering is used to create attributes such as departure hour, day of week, month, route, and flight duration. Exploratory Data Analysis is performed to understand the distribution and frequency of flight delays. Comparative analysis examines delay patterns across airlines, airports, routes, and different time periods. Visualization techniques are used to present delay trends through bar charts, line graphs, heatmaps, and dashboards. Relevant features are selected for Machine Learning model development. A classification model can be trained to predict whether a flight is likely to be delayed, while regression techniques can be used to estimate delay duration where appropriate. The trained model is evaluated using suitable metrics such as accuracy, precision, recall, F1-score, MAE,

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

or RMSE depending on the prediction task. Finally, the system displays analytical findings and prediction results through an interactive user interface.

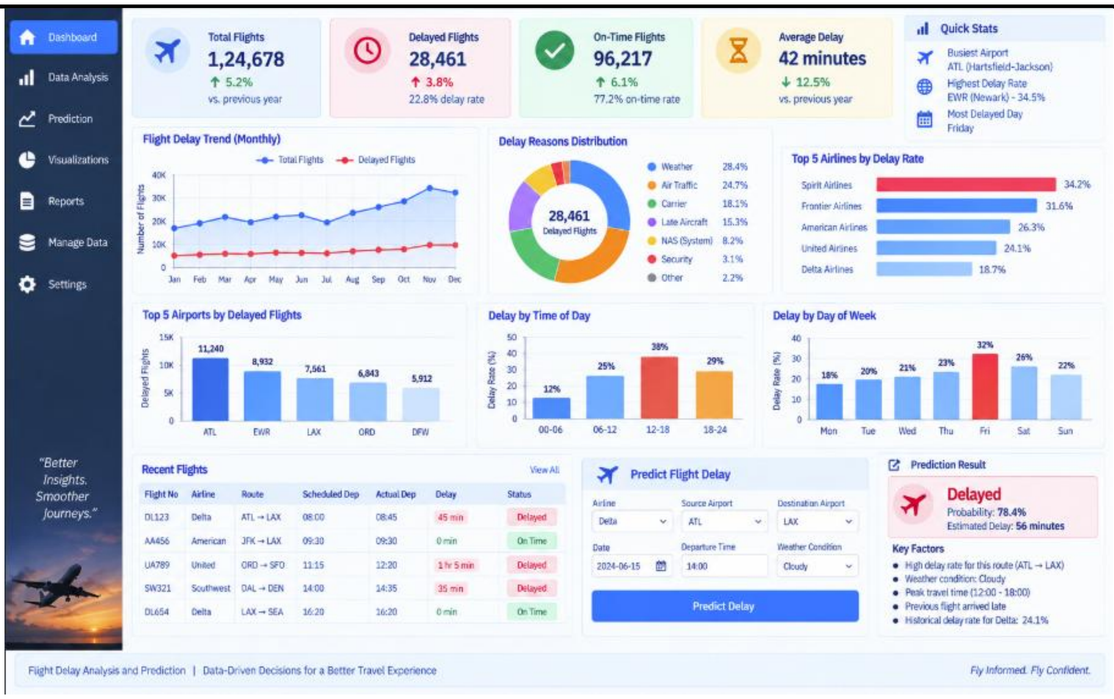
SYSTEM ARCHITECTURE

The system architecture consists of several interconnected layers that process flight data and generate analytical and predictive results. The Data Source Layer provides historical flight records and optional external information such as weather data. The Data Input Layer accepts datasets through CSV, Excel, database, or authorized API sources. The Data Validation Layer checks file formats, required fields, data types, and basic data quality. The Data Preprocessing Layer removes duplicates, handles missing values, standardizes timestamps, and transforms categorical variables. The Data Storage Layer securely stores cleaned flight records and historical information. The Exploratory Analysis Layer calculates flight counts, delay percentages, average delay duration, and other operational metrics. The Delay Pattern Analysis Layer compares delays across airlines, airports, routes, time periods, and other relevant features. The Feature Engineering Layer prepares attributes required for predictive modeling. The Machine Learning Layer trains and evaluates a suitable delay-prediction model using historical data. The Prediction Layer accepts new flight information and produces an estimated delay prediction according to the trained model. The Visualization and Reporting Layer presents trends, comparisons, prediction results, charts, and reports through an interactive dashboard. The User Interface Layer allows authorized users to upload data, explore analytics, view predictions, and generate reports.



V. RESULT AND OUTPUT

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



VI. CONCLUSION

The Flight Delay Analysis and Prediction system provides a data-driven approach for analyzing historical flight information and predicting potential flight delays. The system combines data preprocessing, exploratory data analysis, visualization, and Machine Learning techniques to transform raw flight records into useful insights.

The system analyzes important factors such as airline, origin airport, destination airport, departure time, route, weather conditions, and previous flight information. It identifies patterns in flight delays and provides comparisons across different airlines, airports, routes, days, and time periods.

The analytical dashboard presents important metrics such as total flights, delayed flights, on-time flights, average delay duration, and delay percentage. Charts and graphs make it easier for users to understand delay trends, common delay reasons, airport performance, and time-based patterns.

The Machine Learning component can predict whether a flight may experience a delay based on the available input features. Such predictions can serve as decision-support information for operational planning and passenger awareness. The quality of the prediction depends on the completeness, accuracy, and relevance of the historical data used to train the model.

The proposed system can be useful for airlines, airport analysts, researchers, and transportation-management teams. In the future, it can be enhanced with real-time flight tracking, live weather APIs, advanced prediction models, route-specific forecasting, real-time notifications, and dynamic operational recommendations. These enhancements can make the system more comprehensive and useful for modern aviation data analysis.

REFERENCES

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine

learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, “An advanced movie recommender using collaborative filtering and sentiment analysis,” International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, “Automatic crop recommendation system using LightGBM and decision tree machine learning models,” Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, “Smart agriculture through IoT and machine learning for analyzing carbon footprints,” in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, “Soil image classification using transfer learning approach: MobileNetV2 with CNN,” SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.