International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04786
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Predicting Airline Delays Using Xgboost and Random Forest

[1]CH. VASUNDHARA, [2]DUMMU NAVEEN

1Assistant Professor, 2MCA Final Semester,

[1]Master of Computer Applications,

[1]Sanketika Vidya Parishad Engineering College, Vishakhapatnam, Andhra Pradesh, India

Abstract:
Flight delays pose considerable challenges for both passengers andairlines, resulting in financial burdens, operational inefficiencies, and reduced customer satisfaction. In response to these issues, we present a robust predictive modeling framework that leverages a rich set of historical flight data to accurately forecast potential delays. Our dataset includes features such as scheduled and actual departure times, carrier information, weather conditions, flight routes, and historical delay patterns.We employ the XGBoost regression algorithm, which is well-suited for handling high-dimensional, tabular data and capturing complex nonlinear relationships. Compared to traditional statistical models and baseline machine learning approaches, our XGBoost-based model demonstrates significantly improved predictive accuracy, as measured by metrics such as Mean Absolute Error (MAE) and ROC AUC (for classification-based delay thresholds).A comprehensive feature importance analysis highlights key variables influencing delay likelihood, including departure time of day, weather at origin and destination airports, airline carrier performance, and specific high-traffic routes. These insights not only improve model interpretability but also offer practical guidance for stakeholders in airline scheduling, resourceallocation,and proactive passenger communication.Ultimately, our predictive system enables airlines to anticipate disruptions more effectively, streamline operational planning, reduce costs associated with delays, andenhance the overall passenger experience through more reliable and timely air travel services.

IndexTerms: Machine Learning (ML), Supervised Learning, Semi-supervised Learning, XGBoost, Decision Trees, Random Forest, Logistic Regression

1.INTRODUCTION

In today's fast-paced world, air travel remains a crucial pillar of global connectivity, enabling efficient movement of people and goods. However, a persistent challenge that disrupts this efficiency is flight delays. These delays not only cause inconvenience for travelers but also result in financial losses and logistical complications for airlines and airport authorities. While some delays are inevitable due to unpredictable factors like weather conditions or sudden technical issues, a significant number can be anticipated and managed with the help of data-driven solutions. Recognizing the widespread impact of flight delays, our project focuses on building a real-time flight delay prediction system using advanced machine learning techniques. By leveraging historical flight data, we aim to accurately forecast delays and enhance decision-making for all aviation stakeholders. Our solution is powered by the XGBoost algorithm, known for its performance and accuracy in predictive analytics. What makes our approach unique is the integration of an intelligent recommendation system that suggests alternative flights if the predicted delay exceeds 30 minutes.[12] This added feature not only helps minimize the inconvenience for passengers but also aids in more efficient resource planning for airlines and airports. We have developed a user-friendly web application using Flask that allows users to input flight-specific details such as the date, airline, origin, and 1 destination. The system then delivers real-time predictions regarding potential delays, the estimated duration of the delay, and viable alternative flight options. This interface is designed to simulate real-time usage and can be deployed on cloud platforms like AWS. Additionally, it supports data visualization through dashboards built using tools like Streamlit. By combining predictive modeling with an actionable interface,[18] our project strives to provide a comprehensive solution for proactive flight delay management—empowering travelers, airline operators, and airport authorities with valuable insights and timely alternative.

1.1 Existing System

The existing systems for predicting airline delays predominantly use traditional machine learning algorithms such as Random Forest, Decision Trees, Multi-Layer Perceptron (MLP), and K-Nearest Neighbors (KNN). These models are applied to historical flight data to classify whether a flight will be delayed or not. While they offer basic predictive functionality, these systems often suffer from limitations such as low accuracy, lack of real-time responsiveness, and minimal user interaction capabilities. They typically divide data into training and testing sets for model development but fail to incorporate intelligent features like alternative flight suggestions.[20] Moreover, they rarely integrate deployment strategies or user-friendly web interfaces, making them less practical for real-world use. The models also lack explainability and interpretability, which are crucial for stakeholder trust. Many of these systems do not use real-time data streams or adaptive learning mechanisms, resulting in outdated predictions. Additionally, their inability to update regularly reduces model reliability over time. Another critical drawback is the absence of advanced visualization tools to aid

ISSN: 2583-6129
DOI: 10.55041/ISJEM04786

decision-making. These systems generally overlook cloud deployment or mobile accessibility, limiting user reach. Thus, while existing systems provide a foundational approach to flight delay prediction, they fall short in providing actionable insights, scalability, and interactive features needed for practical deployment in modern aviation contexts.

1.1.1 Challenges:

Data Quality and Availability

- Real-time flight data often contains **missing, inconsistent, or noisy values**, which can affect model accuracy.

Feature Selection and Engineering

- Identifying **relevant and predictive features** from complex flight data is challenging.

Imbalanced Dataset

- The number of **non-delayed flights** may significantly outnumber delayed ones, leading to **class imbalance** and biased predictions.

Real-time Prediction Complexity

- Processing and analyzing **live data streams** in real time is computationally demanding and introduces latency concerns.

Model Overfitting

- Complex models like XGBoost are prone to **overfitting**, especially when hyperparameters aren't tuned properly.

Proposed system:

The proposed system is a real-time flight delay prediction platform built using the XGBoost machine learning algorithm. It analyzes historical flight data such as airline, origin, destination, and scheduled departure time to forecast potential delays. A flight is considered delayed if the departure exceeds the scheduled time by more than 15 minutes. The model is trained on a cleaned and preprocessed dataset, ensuring high prediction accuracy and efficiency. It is integrated into a user-friendly web interface developed using Flask. Users can input flight details like date, month, airline, and route through the web app.[14] Upon submission, the system processes the data and predicts whether the flight will be delayed or on time. If the delay exceeds 30 minutes, the system also suggests alternative flight options. The web application is designed to be accessible and intuitive for passengers, airline staff, and airport authorities. Deployment can be done on cloud platforms like AWS or with visualization dashboards using Streamlit. The system includes features for data visualization, model evaluation, and continuous updates. It supports proactive decision-making by informing users about potential disruptions. The use of XGBoost ensures robustness, scalability, and faster computation. Overall, the system enhances travel planning and reduces the impact of flight delays through intelligent predictions and real-time recommendations.

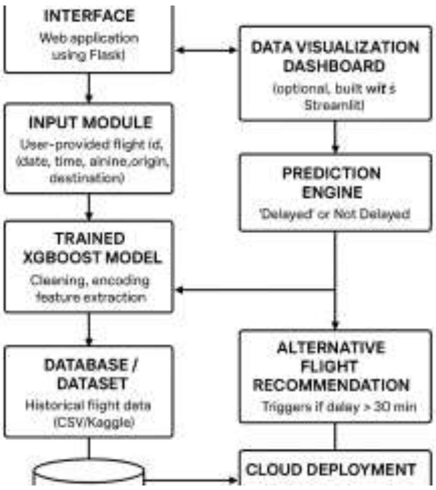

Fig: proposed diagram

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04786
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

UML DIAGRAMS:

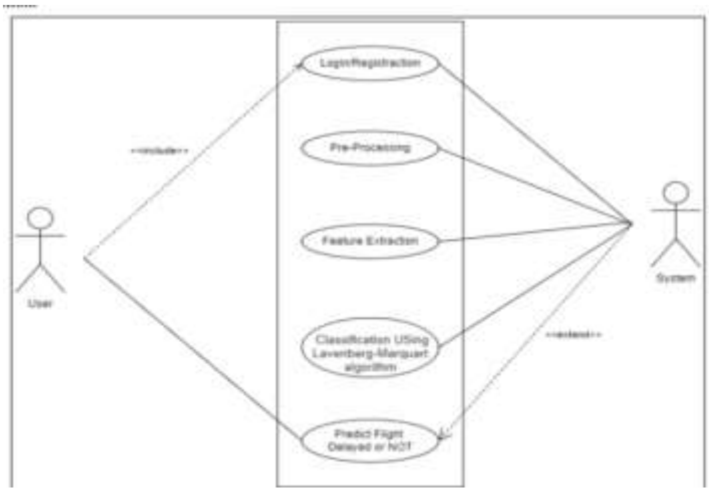

Fig: use case diagram

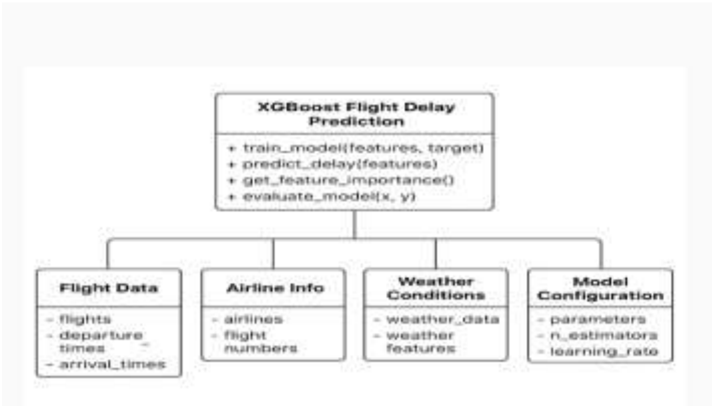

Fig: class diagram

1.1.2 Advantages:

- High Prediction Accuracy

Uses **XGBoost**, a powerful machine learning algorithm known for its high accuracy and efficiency in classification tasks.

- Real-Time Flight Delay Prediction

Users can input flight details and get immediate predictions about potential delays.

- Alternative Flight Suggestions

If a delay exceeds **30 minutes**, the system recommends alternative flights, improving passenger options.

- User-Friendly Web Interface

Built using **Flask**, the web app is easy to use and accessible via browsers for travelers, airline staff, and administrators.

- Data-Driven Insights

Helps understand trends and patterns in delays by analyzing historical flight data.

- Explainable Predictions

Uses feature importance visualization in XGBoost to explain which input factors influenced the prediction.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04786
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Architecture:

The architecture of the flight delay prediction system is designed to enable accurate, real-time delay forecasting using a machine learning model and a web-based interface. At the core of the system is the XGBoost classifier, which is trained on historical flight data, including features such as date, airline, origin, destination, scheduled and actual departure times. The system begins with a user interface (UI) developed using the Flask web framework, where users can input flight-specific details.[8] These inputs are passed to the Input Module, which collects and formats the data. The Data Preprocessing Layer then cleans the inputs, handles missing values, encodes categorical features, and extracts relevant attributes. The cleaned data is fed into the trained XGBoost model via the Prediction Engine, which classifies the flight as either 'Delayed' or 'Not Delayed' based on whether the predicted delay exceeds 15 minutes. If a delay of more than 30 minutes is predicted, the Alternative Flight Recommendation Module is triggered to provide other suitable flight options. Finally, the result—including delay status and alternative suggestions—is displayed back to the user via the web interface. This system is built to support deployment on platforms like AWS and can be integrated with visualization tools like Streamlit to enhance real-time accessibility and usability

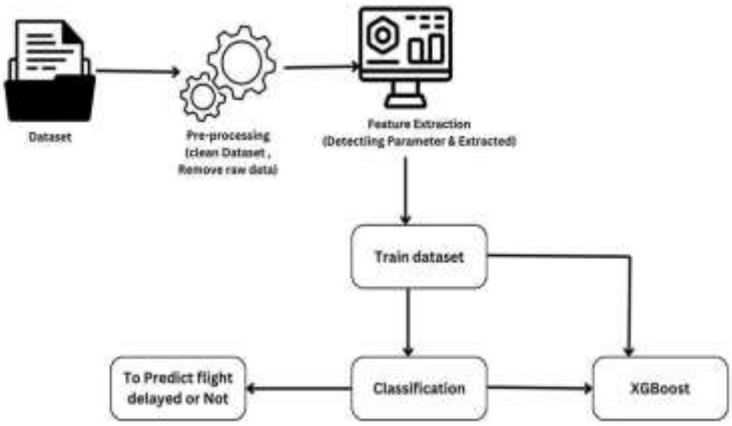

Fig: system architecture

Algorithm:

The primary algorithm used in the project *"Predicting Airline Delays using XGBoost and random forest"* is the XGBoost (Extreme Gradient Boosting) algorithm, which is a powerful and efficient supervised machine learning technique based on gradient boosting.[1] XGBoost is particularly known for its high performance, scalability, and ability to handle large datasets with complex patterns, making it well-suited for flight delay prediction tasks. In this project, the model is trained on historical flight data that includes features such as airline, origin, destination, date, and departure times. The trained XGBoost model is then used to classify whether a given flight is likely to be delayed or not, based on a threshold of 15 minutes. If the delay exceeds 30 minutes, the system also triggers a recommendation for alternative flights. XGBoost's ability to minimize overfitting, manage missing data, and deliver fast training and prediction times made it the ideal choice for implementing this real-time delay prediction system. Although other algorithms like Random Forest, Decision Trees, and Support Vector Machines were explored in the literature review, XGBoost was ultimately selected for its superior accuracy, flexibility, and robust performance in handling classification problems in aviation data.

Techniques:

The project *"Predicting Airline Delays using XGBoost and random forest"* utilizes several key techniques from machine learning and software development to achieve accurate flight delay prediction and an interactive user experience. The central technique is supervised learning, specifically classification, where the model learns from labeled historical flight data to predict whether a flight will be delayed or not. The project uses data preprocessing techniques such as handling missing values, encoding categorical variables (like airlines and airport codes), feature selection, and normalization to prepare the dataset for model training. Feature engineering is applied to derive meaningful features, such as calculating departure delay durations, which improve the model's predictive capability.[5] The core machine learning technique used is the XGBoost algorithm, a form of gradient boosting, chosen for its accuracy, speed, and regularization features that reduce overfitting. The trained model is integrated into a web-based application using Flask, a lightweight Python web framework, allowing users to input flight details and receive real-time predictions. Additionally, data visualization techniques using libraries like Matplotlib and Seaborn are employed to display trends in flight delays and feature importance.[10] The project also incorporates alternative flight recommendation logic if predicted delays exceed 30 minutes, enhancing its usability. Together, these techniques create a robust, scalable, and user-friendly solution for proactive flight delay management.

Tools:

The project *"Predicting Airline using XGBoost and random forest"* employs a variety of software tools and technologies to build a functional and efficient flight delay prediction system. The main programming language used is Python, chosen for its simplicity and rich ecosystem of data science libraries. For data analysis and manipulation, the project utilizes Pandas and NumPy, which help in handling large datasets and performing numerical computations. Seaborn and Matplotlib are used for data visualization, allowing developers and users to understand delay patterns and feature importance visually.[7] The XGBoost library is the core machine learning tool, providing the algorithm used for training and predicting flight delays. To develop the web interface, the project uses Flask, a lightweight web framework in Python that facilitates user interaction with the machine learning model through a browser-based form. The Visual Studio Code (VS Code) IDE is used for writing and debugging code efficiently. For saving and loading the trained model, Pickle is employed for model serialization. Additionally, the project supports deployment on cloud platforms like AWS and can be extended to dashboards using Streamlit.[9] These tools collectively ensure that the system is not only accurate and reliable but also accessible, interactive, and scalable for real-world use.

Methods:

The project *"Predicting Airline Delays using XGBoost and random forest"* employs several key methods throughout its development and implementation to ensure accurate predictions and user-friendly interaction.[17] The first method used is data collection, where historical flight data is gathered from open sources like Kaggle, containing attributes such as flight date, airline, origin, destination, departure time, and delay duration. This is followed by data preprocessing methods, which include cleaning the dataset, handling missing values, converting time fields to a usable format, and encoding categorical variables into numerical form using label encoding. The feature engineering method is applied to create new variables like 'departure delay' and to filter out irrelevant features, enhancing the model's predictive power. For model training, the train-test split method is used to divide the data into training and testing sets, ensuring the model is validated before deployment.[13] The project then uses the XGBoost classification method to learn from the training data and predict whether a flight will be delayed. The model's performance is evaluated using standard methods like accuracy scoring and visualization of feature importance. Finally, the Flask web framework method is used to develop an interactive web interface, allowing users to input flight details and receive real-time predictions. If a flight is predicted to be delayed beyond 30 minutes, the system uses a conditional method to trigger alternative flight suggestions, making the solution both intelligent and user-centric.

METHODOLOGY

Input:

In the project *"Predicting Airline Delays using XGBoost and random forest"*, the input information consists of specific flight-related details provided by the user through a web-based interface. These inputs are crucial features required by the machine learning model to make an accurate delay prediction. The user is prompted to enter the year, month, and day of the flight, as well as the airline carrier, origin airport, and destination airport. Additional attributes such as the scheduled departure time and the actual departure time are also part of the dataset used during model training, though not all are manually input by users at prediction time. These inputs are then preprocessed—categorical fields like airline and airport codes are label encoded, and numerical fields are scaled or transformed as needed—to match the format expected by the trained XGBoost model. [15] Once the input data is processed, it is fed into the model, which then predicts whether the flight is likely to be delayed. If the expected delay exceeds 30 minutes, the system also recommends alternative flights, enhancing the decision-making capability for the user.

ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025
DOI: 10.55041/ISJEM04786
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

```
flights_needed_data.info()  # for an insight into the data
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 31 columns):
 #   Column               Non-Null Count   Dtype
---  ------               --------------   -----
 0   YEAR                 100000 non-null  int64
 1   MONTH                100000 non-null  int64
 2   DAY                  100000 non-null  int64
 3   DAY_OF_WEEK          100000 non-null  int64
 4   AIRLINE              100000 non-null  object
 5   FLIGHT_NUMBER        100000 non-null  int64
 6   TAIL_NUMBER          99833 non-null   object
 7   ORIGIN_AIRPORT       100000 non-null  object
 8   DESTINATION_AIRPORT  100000 non-null  object
 9   SCHEDULED_DEPARTURE  100000 non-null  int64
 10  DEPARTURE_TIME       97702 non-null   float64
 11  DEPARTURE_DELAY      97702 non-null   float64
 12  TAXI_OUT             97629 non-null   float64
 13  WHEELS_OFF           97629 non-null   float64
 14  SCHEDULED_TIME       100000 non-null  float64
 15  ELAPSED_TIME         97387 non-null   float64
 16  AIR_TIME             97387 non-null   float64
 17  DISTANCE             100000 non-null  int64
 18  WHEELS_ON            97560 non-null   float64
 19  TAXI_IN              97560 non-null   float64
...
 29  LATE_AIRCRAFT_DELAY  34625 non-null   float64
 30  WEATHER_DELAY        34625 non-null   float64
dtypes: float64(16), int64(10), object(5)
memory usage: 23.7+ MB
```

Fig: Flight Data

Method of Process:

In the project *"Predicting Airline Delays using XGBoost and random forest"*, the method of processing information follows a structured pipeline that transforms raw user input into actionable predictions. The process begins with data input collected via a web interface, where users provide flight-specific details such as date, airline, origin, and destination. This input is passed to the data preprocessing module, which applies several key methods: it handles missing values, encodes categorical variables using label encoding, and extracts meaningful features like departure delay duration.[2] After preprocessing, the cleaned and structured data is forwarded to the trained XGBoost model, which has been previously trained on historical flight data. [11] The model then applies its learned patterns to classify whether the flight will be delayed or not, based on a defined threshold (e.g., 15 minutes). If a delay longer than 30 minutes is predicted, the system further triggers the alternative flight recommendation method, providing users with backup options. Finally, the result—either "Delayed" or "Not Delayed," along with suggested alternatives—is displayed back to the user through the interface. This end-to-end process ensures that user-provided information is systematically analyzed and transformed into accurate, real-time flight delay predictions.

Output:

In the project *"Predicting Airline Delays using XGBoost and random forest"*, the output information is the final result generated by the trained machine learning model after processing the user-provided flight details.[6] The primary output is a classification result indicating whether the flight is predicted to be "Delayed" or "Not Delayed", based on a threshold of 15 minutes difference between scheduled and actual departure times. In addition to this binary outcome, the system also provides an estimated delay duration, giving users a clearer idea of how long the delay might be. Furthermore, if the predicted delay exceeds 30 minutes, the system triggers an alternative flight recommendation, offering users practical options to minimize disruption.[19] This output is displayed in a user-friendly manner through a Flask-based web interface, making it easy for travelers to interpret the results. The output also includes performance metrics like model accuracy (e.g., 88%) during evaluation, and visual outputs such as feature importance graphs and delay distribution plots using tools like Matplotlib and Seaborn. Together, these outputs provide comprehensive, real-time, and actionable information to support informed travel planning.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04786
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

```python
# filtering out unnecessary columns
flights_needed_data=flights_needed_data.drop(['YEAR','FLIGHT_NUMBER','AIRLINE','DISTANCE','TAIL_NUMBER','TAXI_OUT',
                                              'SCHEDULED_TIME','DEPARTURE_TIME','WHEELS_OFF','ELAPSED_TIME',
                                              'AIR_TIME','WHEELS_ON','DAY_OF_WEEK','TAXI_IN','CANCELLATION_REASON'],
                                             axis=1)
```

```python
flights_needed_data
```

	MONTH	DAY	ORIGIN_AIRPORT	DESTINATION_AIRPORT	SCHEDULED_DEPARTURE	DEPARTURE_DELAY	SCHEDULED_ARRIVAL	ARRIVAL_TIME	ARRIVAL_DELAY	DIVERTED	CANCELLED
0	1	1	ANC	SEA	5	-11.0	430	408.0	-22.0	0	0
1	1	1	LAX	PBI	10	-8.0	750	741.0	-9.0	0	0
2	1	1	SFO	CLT	20	-2.0	806	811.0	5.0	0	0
3	1	1	LAX	MIA	20	-5.0	805	756.0	-9.0	0	0
4	1	1	SEA	ANC	25	-1.0	320	259.0	-21.0	0	0
...	...	...	...	...	...	...	...	...	...	...	...
99995	1	7	ATL	BQK	1108	-6.0	1219	1208.0	-11.0	0	0
99996	1	7	LAS	PHL	1108	9.0	1842	1844.0	2.0	0	0
99997	1	7	SFO	BFL	1108	-7.0	1225	1232.0	7.0	0	0
99998	1	7	ORD	MCO	1109	7.0	1454	1453.0	-1.0	0	0
99999	1	7	HOU	DFW	1109	-9.0	1220	1212.0	-8.0	0	0

100000 rows × 16 columns

Fig: Flight Needed Data

```python
# removing some more columns
flights_needed_data=flights_needed_data.drop(['ORIGIN_AIRPORT', 'DESTINATION_AIRPORT', 'ARRIVAL_TIME', 'ARRIVAL_DELAY'],axis=1)
flights_needed_data
```

	MONTH	DAY	SCHEDULED_DEPARTURE	DEPARTURE_DELAY	SCHEDULED_ARRIVAL	DIVERTED	CANCELLED	AIR_SYSTEM_DELAY	SECURITY_DELAY	AIRLINE_DELAY	LATE_AIRCRAFT_DELA
0	1	1	5	-11.0	430	0	0	NaN	NaN	NaN	Na
1	1	1	10	-8.0	750	0	0	NaN	NaN	NaN	Na
2	1	1	20	-2.0	806	0	0	NaN	NaN	NaN	Na
3	1	1	20	-5.0	805	0	0	NaN	NaN	NaN	Na
4	1	1	25	-1.0	320	0	0	NaN	NaN	NaN	Na
...	...	...	...	...	...	...	...	...	...	...	
99995	1	7	1108	-6.0	1219	0	0	NaN	NaN	NaN	Na
99996	1	7	1108	9.0	1842	0	0	NaN	NaN	NaN	Na
99997	1	7	1108	-7.0	1225	0	0	NaN	NaN	NaN	Na
99998	1	7	1109	7.0	1454	0	0	NaN	NaN	NaN	Na
99999	1	7	1109	-9.0	1220	0	0	NaN	NaN	NaN	Na

100000 rows × 13 columns

Fig: Flight history

```python
sb.heatmap(corr)
```

```
<Axes: >
```

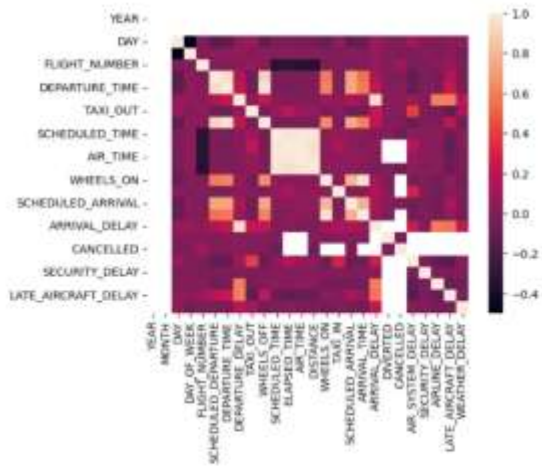

Fig:Heatmap

ISSN: 2583-6129
DOI: 10.55041/ISJEM04786

RESULTS:

The result of the project *"Predicting Airline Delays using XGBoost and random forest"* is a real-time flight delay prediction system that uses a trained XGBoost model with 88% accuracy to forecast delays based on flight details like date, airline, origin, and destination. Integrated into a Flask-based web app, it provides instant predictions and suggests alternative flights if delays exceed 30 minutes.[16] The system also includes visualizations for better insight, offering a practical and user-friendly tool for passengers and airline operators.

DISCUSSIONS:

The project *"Predicting Airline Delays using XGBoost and random forest"* includes several important discussions focused on improving the accuracy and usability of flight delay prediction systems. It highlights the limitations of traditional methods and existing machine learning approaches, such as low accuracy, lack of real-time data usage, and absence of user-friendly interfaces or alternative flight suggestions. To overcome these issues, the project proposes a robust solution using the XGBoost algorithm, known for its speed and precision in classification tasks. The discussion emphasizes the importance of data preprocessing, feature engineering, and model evaluation in building a reliable predictive system. It also covers the integration of the model into a Flask web application for real-time user interaction and introduces a unique feature—alternative flight recommendations for delays over 30 minutes.[4] Additionally, the project discusses future improvements, such as including weather and live traffic data, enhancing model interpretability, and deploying the system on cloud platforms. These discussions demonstrate a comprehensive understanding of both the technical and practical challenges in flight delay prediction and offer a clear path for future enhancements.

CONCLUSION:

The conclusion of the project *"Predicting Airline Delays using XGBoost and random forest"* highlights the successful development of an intelligent, real-time flight delay prediction system that effectively addresses one of the most common challenges in the aviation industry. By leveraging the power of the XGBoost algorithm and historical flight data, the system accurately predicts whether a flight will be delayed, achieving an accuracy of 88%. Integrated with a user-friendly Flask web application, the model allows users to input flight details and receive instant predictions along with alternative flight suggestions if the delay exceeds 30 minutes. This project not only enhances passenger experience by reducing uncertainty and enabling better planning but also supports airlines and airport authorities in improving operational efficiency. Overall, the system demonstrates how machine learning can be applied to solve real-world problems and serves as a strong foundation for future enhancements such as incorporating real-time weather data and expanding functionality.

FUTURE SCOPE:

The future scope of the project includes integrating real-time data such as weather, air traffic, and maintenance logs to enhance prediction accuracy. It can also incorporate tools like SHAP or LIME for better model interpretability. Additional features like interactive dashboards, user alerts, and mobile support can improve usability. The system can be extended to assist with flight scheduling and airport operations, and deployed on cloud platforms with continuous updates for real-world use.

ACKNOWLEDGEMENT:

Chinthagingala Vasundhara: working as an Assistant professor in master of computer application in sanketika vidya parishad engineering college, Visakhapatnam Andhra Pradesh. With 2 years of experience in computer science and engineering (CSE), accredited by NAAC.with her area of intrest in java full stack.She is dedicated and emerging academician in the field of Computer Science, currently serving as a faculty member. With a strong foundation in technical concepts and a passion for teaching, she has begun his academic journey by actively mentoring students in practical and innovative projects. As a faculty, she has already made a significant impact by guiding student teams through their academic projects with clarity, enthusiasm, and technical proficiency. His commitment to student success and interest in research-driven teaching make him a promising contributor to academic excellence

DUMMU NAVEEN is pursuing his final semester MCA in Sanketika Vidya Parishad Engineering College, accredited with A grade by NAAC, affiliated by Andhra University and approved by AICTE. With interest in Machine learning Dummu Naveen has taken up her PG project on PREDICTING AIRLINE DELAYS USING XG BOOST AND RANDOM FORESTand published the paper in connection to the project under the guidance of CH. VASUNDHARA, Assistant Professor, SVPEC.

REFRENCES:

[1] Predicting Flight Delays with Machine Learning: A Case Study from Saudi Arabian Airlines

https://onlinelibrary.wiley.com/doi/full/10.1155/2024/3385463

[2] Prediction of Flight Delays Based on the XGboost Model

https://ieeexplore.ieee.org/abstract/document/10788297

[3] Flight delay prediction based with machine learning

https://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-68c7e1b9-e46c-457d-8e8a-d477926c07cd

[4] Flight Delay Prediction Using Gradient Boosting Machine Learning Classifiers

https://www.proquest.com/openview/d5919d233c46f69e5c7f99ed72f19718/1?pq-origsite=gscholar&cbl=4585454

[5] Predictive Modeling of Flight Delays at an Airport Using Machine Learning Methods

https://www.mdpi.com/2076-3417/14/13/5472

[6] Flight Arrival Delay Prediction Using Gradient Boosting Classifier

https://link.springer.com/chapter/10.1007/978-981-13-1498-8_57

[7] Machine Learning Model - based Prediction of Flight Delay

https://ieeexplore.ieee.org/abstract/document/9243339

[8] XGBoost for Predicting Airline Customer Satisfaction Based on Computational Efficient Questionnaire

https://socjs.telkomuniversity.ac.id/ojs/index.php/ijoict/article/view/864

[9] XGBoost in Public Transportation for Multi-Attribute Data: Delay Prediction in Railway Systems in Real-Time

https://ieeexplore.ieee.org/abstract/document/10683692

[10] Flight Delay Prediction Based on Gradient Boosting Ensemble Techniques

https://ieeexplore.ieee.org/abstract/document/10016828

[11] Flight Delay Prediction for Mitigation of Airport Commercial Revenue Losses Using Machine Learning on Imbalanced Dataset

https://ieeexplore.ieee.org/abstract/document/10037369

[12] Flight Delay Prediction Using Machine Learning

https://ieeexplore.ieee.org/abstract/document/10126220

ISSN: 2583-6129
DOI: 10.55041/ISJEM04786

[13] Flight Arrival Delay Prediction And Analysis Using Ensemble Learning

https://ieeexplore.ieee.org/abstract/document/9084929

[14] Flight Arrival Delay Prediction Using Supervised Machine Learning Algorithms

https://link.springer.com/chapter/10.1007/978-3-030-72588-4_16

[15] Predicting the likelihood of airspace user rerouting to mitigate air traffic flow management delay

https://www.sciencedirect.com/science/article/abs/pii/S0968090X22002844

[16] Flight delay prediction based on LightGBM

https://ieeexplore.ieee.org/abstract/document/9633431

[17] Research on Flight Delay Prediction Based on Horizontal and Vertical Federated Learning Framework

https://ieeexplore.ieee.org/abstract/document/9633398

[18] Predicting flight delay with spatio-temporal trajectory convolutional network and airport situational awareness map

https://www.sciencedirect.com/science/article/abs/pii/S092523122101612X

[19] Flight Delay Prediction Using Different Regression Algorithms in Machine Learning

https://ieeexplore.ieee.org/abstract/document/10125675

[20] Considering Airport Planners' Preferences and Imbalanced Datasets when Predicting Flight Delays and Cancellations

https://ieeexplore.ieee.org/abstract/document/9594367

