VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

(An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering)

Department of Computer Engineering

Project Report on

(JetLagged - Prediction of Airline Flight Delay)

Submitted in partial fulfillment of the requirements of Third Year (Semester–VI), Bachelor of Engineering Degree in Computer Engineering at the University of Mumbai Academic Year 2024-25

By

Ved Waje (64)
Abhirat More (44)
Pranita Bannore (08)
Harshita Lohana (38)

Project Mentor
Mrs. Vidya Zope

University of Mumbai
(AY 2024-25)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

(An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering)

Department of Computer Engineering

CERTIFICATE

This is to certify that ____ of Third Year Computer Engineering studying under the University of Mumbai has satisfactorily presented the project on **"JetLagged-Prediction of Airline Flight Delay"** as a part of the coursework of Mini Project 2B for Semester-VI under the guidance of **Mrs. Vidya Zope** in the year 2024-25.

____

Date

____________________ ____________________

Internal Examiner External Examiner

____________________ ____________________ ____________________

Project Mentor	Head of the Department	Principal
	Dr. Mrs. Nupur Giri	Dr. J. M. Nair

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea / data / fact / source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

--	--
(Signature)	(Signature)
--	--
Ved Waje (64)	Abhirat More (44)
--	--
(Signature)	(Signature)
--	--
Pranita Bannore (08)	Harshita Lohana (38)

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Dr. (Mrs.) Priya R. L** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department

COURSE OUTCOMES FOR T.E MINI PROJECT 2B

Learners will be to:-

CO No.	COURSE OUTCOME
CO1	Identify problems based on societal /research needs.
CO2	Apply Knowledge and skill to solve societal problems in a group.
CO3	Develop interpersonal skills to work as a member of a group or leader.
CO4	Draw the proper inferences from available results through theoretical/ experimental/simulations.
CO5	Analyze the impact of solutions in societal and environmental context for sustainable development.
CO6	Use standard norms of engineering practices
CO7	Excel in written and oral communication.
CO8	Demonstrate capabilities of self-learning in a group, which leads to lifelong learning.
CO9	Demonstrate project management principles during project work.

ABSTRACT

In the present scenario of airline flights, there have been numerous instances of flight delays and cancellations. Flight delays and cancellations are common issues in the airline industry, affecting both passengers and airlines. These disruptions can result from a variety of factors, including adverse weather conditions, technical malfunctions, air traffic congestion, and crew availability problems. To tackle these challenges, this project aims to develop an advanced deep learning based web application designed to identify and analyze the primary causes of flight delays and cancellations. This application will integrate and analyze data from multiple sources, such as weather reports, flight schedules, and historical delay records. Notably, the interconnected nature of these factors poses a significant challenge, as delays in one segment can lead to cascading effects throughout the entire network. By employing sophisticated deep learning models, it will detect patterns and correlations that contribute to these disruptions. The integration of such diverse datasets allows for a comprehensive understanding of the dynamics involved in flight delays. The primary objective is to offer a comprehensive tool that enables airlines and passengers to understand and predict potential delays more effectively. Additionally, the application aims to provide real-time alerts and recommendations to the passengers. This proactive approach will facilitate better decision-making, thereby enhancing overall operational efficiency within the airline industry. Furthermore, by leveraging advanced data analytics and deep learning techniques, the system seeks to contribute to a more resilient airline industry, potentially paving the way for future innovations in predictive analytics.

The proposed system will be an invaluable resource for analyzing and addressing the factors that lead to flight delays. By improving the ability to predict and manage these disruptions, the system aims to enhance the travel experience for passengers and reduce operational costs for airlines. Ultimately, this project represents a significant advancement in utilizing deep learning technology to improve reliability and efficiency in air travel, benefiting the entire airline ecosystem through innovative and data-driven solutions. This initiative aligns with the broader goals of digital transformation and smart transportation systems, highlighting the role of technology in shaping the future of global connectivity.

Index

Title **page no.**

1. Introduction

1.1 Introduction

The airline industry plays a crucial role in global transportation, facilitating the movement of millions of passengers and goods daily. However, flight delays and cancellations are persistent issues that disrupt travel schedules, cause financial losses, and lead to passenger dissatisfaction. Understanding the underlying causes of these disruptions is essential for improving airline operations and enhancing the passenger experience. Flight delays and cancellations can stem from various factors, including adverse weather conditions, technical malfunctions, late arrivals of connecting flights, and crew-related issues. These factors can be interrelated, making it challenging to predict and manage delays. Traditional methods of delay prediction often fall short due to the complexity and volume of data involved. In this project, we propose the development of a deep learning-based web application to identify and analyze the primary causes of flight delays and cancellations. By leveraging advanced deep learning techniques, the application aims to process and analyze large datasets, including weather data, flight schedules, and historical delay records. The goal is to uncover patterns and correlations that contribute to flight disruptions, providing valuable insights for both airlines and passengers. The application will not only predict potential delays but also offer recommendations for mitigating their impact. By providing real-time data and predictive insights, the system can assist airlines in optimizing their operations and improving decision-making processes. Passengers, on the other hand, can benefit from timely information about potential delays, allowing them to plan their journeys more effectively.

1.2 Motivation

The airline industry plays a pivotal role in the global transportation network, enabling the efficient movement of passengers and goods across the world. Every day, millions of people rely on airlines to connect cities, countries, and continents, making air travel an indispensable part of modern life. However, one of the most persistent and disruptive challenges faced by both airlines and passengers is flight delays and cancellations. These disruptions not only inconvenience travelers but also lead to substantial economic losses for airlines and other related industries. As air traffic continues to increase, there is a growing need to find more effective ways to predict and manage these delays to enhance operational efficiency and improve the overall passenger experience. the potential impact this project could have on both airlines and passengers. For airlines, more accurate delay predictions can lead to better resource allocation, improved decision-making, and reduced operational costs. Airlines can proactively adjust their schedules, manage crew availability, and optimize airport operations based on real-time predictions, resulting in fewer delays and improved overall efficiency. For passengers, access to real-time delay information allows for better travel planning, reducing stress and uncertainty associated with unexpected disruptions. Passengers can make informed decisions about rebooking, adjusting connections, or altering travel plans, leading to a smoother and more pleasant travel experience. The motivation behind this project is driven by the desire to solve a pressing problem in the airline industry using deep learning techniques. By providing airlines and passengers with accurate, real-time insights into potential disruptions, we hope to make air travel more reliable, efficient, and enjoyable for everyone involved. The potential for widespread application and long-term benefits makes this project both timely and impactful, with the ability to significantly improve the way we approach and manage flight delays in the future.

1.3 Problem Statement & Objectives

The airline industry is frequently challenged by flight delays and cancellations, which can result in significant economic losses and inconvenience for both airlines and passengers. These disruptions can be caused by a multitude of factors, including but not limited to:

1. **Adverse Weather Conditions**: Weather events such as thunderstorms, snowstorms, and fog can lead to delays and cancellations, disrupting flight schedules and causing logistical challenges.
2. **Technical Malfunctions**: Mechanical issues or system failures in aircraft can lead to delays while repairs or maintenance are conducted to ensure passenger safety.
3. **Air Traffic Congestion**: High traffic volumes at airports can cause delays in takeoffs and landings, leading to cascading delays across the network.
4. **Crew-Related Issues**: Crew availability, regulatory rest requirements, and logistical challenges in crew scheduling can also contribute to delays.
5. **Operational Inefficiencies**: Inefficient processes in ground handling, baggage handling, and other operational aspects can result in delays.

Despite the availability of data from various sources, including weather reports, flight schedules, and historical flight data, the airline industry lacks an integrated, data-driven solution to predict and understand the root causes of these delays and cancellations effectively. Traditional predictive models often fail to capture the complex interplay of factors that lead to disruptions.

To address this gap, the project aims to develop a deep learning-based web application that can analyze large datasets and identify the primary causes of flight delays and cancellations. By utilizing advanced deep learning techniques, the application will provide real-time predictions and insights, helping airlines optimize their operations and improve decision-making. Additionally, the application will inform passengers about potential delays, enhancing their travel planning and experience.

1.4 Organization of the Report

The **Literature Survey** section provides a comprehensive review of existing systems and research related to jet lag, flight delay prediction, and other relevant topics. This part of the report discusses the current technologies, tools, and methodologies used in the field. It highlights their capabilities and limitations, establishing a foundation for understanding the current state of the art. The **Survey of Existing Systems** dives deeper into the mechanisms behind these systems, showcasing their strengths and areas where they fall short. Following this, the **Limitations of Existing Systems / Research Gap** outlines the critical issues or inefficiencies that remain unaddressed, such as limited predictive accuracy, outdated data sources, or lack of user-centered design. This gap paves the way for our contribution. The **Mini Project Contribution** section presents the unique aspects of our project, *Jetlagged*, by demonstrating how it fills these gaps. It could involve innovations in model accuracy, better data integration (such as live flight tracking), or improved user experience through the mobile app interface.

In the **Proposed System** section, the report begins with an **Introduction** to our system, explaining its purpose and significance. This is followed by the **Architectural Framework / Conceptual Design**, where the overall structure and key components of *Jetlagged* are detailed, explaining how data flows from input (user flight details, API data) to output (predicted delays and suggestions). The **Algorithm and Process Design** delves into the specific algorithms and processes used for predictions, such as any machine learning models, data preprocessing steps, or decision rules based on weather and flight conditions. Next, the **Methodology Applied** section elaborates on the overall approach, including the steps taken for system development, testing, and deployment. The **Hardware and Software Specifications** outlines the technical requirements, such as tools, libraries, and platforms used in both development and execution phases.

The **Experiment and Results for Validation and Verification** section presents the results from testing the system, showcasing how well the proposed solution works in real-world conditions. It includes experiments to verify the accuracy of the predictions and performance metrics compared to existing solutions. In **Result Analysis and Discussion**, the findings from the experiments are analyzed in detail. This includes a discussion of the strengths, any unexpected results, and how they compare to the original goals of the project. Finally, the **Conclusion and Future Work** section wraps up the report by summarizing the key outcomes and suggesting possible future improvements, such as expanding the system to incorporate additional data sources or further optimizing the prediction algorithms.

1.5 Lacuna of the existing systems

The aviation industry continues to face several challenges that hinder its efficiency and overall performance. One major issue is the inadequate predictive accuracy of existing systems, which limits the ability to anticipate delays, disruptions, and maintenance needs. Additionally, there is often a limited integration of data sources, preventing a holistic view of operations and passenger behavior. This contributes to a lack of real-time insights, making it difficult for airlines to respond swiftly to changing circumstances. Moreover, the insufficient analysis of interrelated factors—such as weather, staffing, and aircraft status—further weakens decision-making. As a result, airlines tend to take reactive rather than proactive measures, which compromises efficiency and reliability. Passengers are also impacted by a poor user experience, marked by unclear communication and delays. Ultimately, these issues lead to suboptimal airline operations, affecting profitability and customer satisfaction alike.

1.6 Relevance of the Project

The relevance of this project lies in its ability to address one of the most pressing challenges in the aviation industry—flight delays and cancellations. These disruptions not only cause inconvenience to passengers but also result in significant financial losses and operational inefficiencies for airlines. Traditional methods of delay prediction often struggle to process and interpret the vast and complex datasets involved, such as weather conditions, technical issues, air traffic congestion, and crew availability. This project introduces a deep learning-based web application designed to integrate and analyze these diverse data sources in real time.

By uncovering hidden patterns and correlations, the system can predict potential delays more accurately and provide actionable insights. This helps airlines improve scheduling, optimize resource allocation, and make informed decisions quickly. For passengers, the application offers timely updates and recommendations, enhancing their ability to plan ahead and reducing the frustration caused by unexpected delays.

Moreover, the project supports the broader goal of building smarter and more resilient transportation systems. As the aviation industry increasingly embraces digital transformation, such intelligent systems are essential for boosting reliability and efficiency. This project thus plays a critical role in advancing predictive analytics in air travel, making it more adaptive, data-driven, and passenger-centric.

2. Literature Survey

A. Overview of Literature Survey

The increasing demand for reliable and efficient air travel has made flight delay prediction a critical area of research. Accurate prediction not only enhances operational planning for airlines but also improves the passenger experience by minimizing uncertainty and inconvenience. Over the years, researchers have explored various methodologies—ranging from statistical techniques and machine learning models to more recent deep learning approaches—to address the problem of flight delays. This literature survey aims to provide a comprehensive understanding of the existing research, identify their strengths and limitations, and establish the need for a more integrated and dynamic solution. The insights gained from this survey serve as the foundation for developing an improved delay prediction model that leverages real-time data and advanced algorithms.

B. Related Works

Several research papers have contributed to the field of flight delay prediction. Early studies primarily relied on historical data and statistical models, which, while useful, struggled to accommodate the nonlinear and dynamic nature of flight operations. More recent works have implemented machine learning algorithms such as Support Vector Machines (SVM), XGBoost, and Multi-Layer Perceptrons (MLP), demonstrating improved accuracy. Some studies have gone further, adopting advanced deep learning techniques such as Graph Convolutional Networks (GCN) and Dynamic Temporal Convolutional Networks (DTCN), which can model complex relationships and time-based dependencies in the data. Despite these advancements, most models still lack the integration of real-time data and comprehensive variable analysis, highlighting the need for a more holistic and proactive approach to flight delay prediction.

2.1 Research Papers Referred

Paper Title	Description
A Deep Learning Approach for Flight Delay Prediction Through Time-Evolving Graphs [1]	Uses GCN to model time-varying spatial interactions in multi-airport networks for delay prediction
Flight Delay Prediction Using Machine Learning [2]	Compares ML algorithms for predicting weather-related delays at JFK airport, with XGBoost performing best
Iterative Machine and Deep Learning Approach for Aviation Delay Prediction [3]	Employs iterative refinement process with advanced algorithms to improve delay prediction accuracy
A Flight Delay Prediction Model with Dynamic Temporal Convolutional Network [4]	Leverages DTCN to capture complex temporal dependencies in flight data for enhanced prediction performance
Flight Delay Prediction Based on Multi-Layer Perceptron [5]	Demonstrates effectiveness of neural networks in handling complexity of flight delay factors
Research on Flight Arrival Delay Prediction Based on Support Vector Machine [6]	Uses SVM with flight plan data, highlighting impact of planning and sample size on prediction accuracy
A Review of Flight Delay Prediction Methods [7]	Reviews statistical, simulation, queuing theory, and ML approaches for delay prediction
Machine Learning Model-Based Prediction of Flight Delay [8]	Applies XGBoost for binary classification and linear regression to predict arrival delays

Table 2.1: description of research papers referred

2.2 Patent search

To understand the innovation landscape, a search was conducted for patents related to flight delay prediction and management systems. Several patents were found focusing on: Notification systems for communicating delays to passengers. Centralized systems for managing airline operations based on estimated times of arrival. Weather-based delay analysis systems.

However, these patented systems generally rely on static rule-based frameworks or basic machine learning models. They do not employ deep learning methods or integrate diverse data sources such as live air traffic, maintenance logs, or real-time weather updates. Furthermore, many focus on post-delay communication rather than proactive prediction and prevention of delays.

2.3. Inference drawn

Existing systems are largely reactive, focusing on alerting stakeholders after delays occur. Few systems utilize real-time data streams or modern deep learning models capable of learning complex patterns. There is a significant gap in end-to-end systems that can predict, visualize, and suggest operational actions in real time. Current patents lack comprehensive data integration—rarely combining weather, air traffic control, crew schedules, and flight connections into a single predictive model. These findings reinforce the need for a new solution that is not only predictive but also actionable, leveraging real-time data and advanced analytics.

2.4 Comparison with the existing system

Parameter	Existing Systems	Proposed Mini Project
Prediction Model	Statistical/Basic ML models	Deep Learning models (e.g., MLP, DTCN)
Real-Time Data Usage	Limited or None	Incorporates real-time weather, traffic, and flight status
Data Sources	Single or few sources (e.g., weather, schedule)	Multi-source: weather, air traffic, crew, maintenance
Scope of Prediction	Focus on specific delays (e.g., weather-related)	Holistic prediction considering multiple factors
Proactivity	Reactive (delay notification)	Proactive (delay prediction and mitigation suggestions)
Interface for Users	Often backend or admin-only tools	User-facing Android and web app with actionable insights
Integration of Cascading Effects	Mostly ignored	Accounted for (e.g., delays at one airport affecting others)
Adaptability	Static, based on historical data	Adaptive through live updates and dynamic learning

Table 2.2: comparison of project with existing system

3. Requirement Gathering for the Proposed System

3.1 Introduction to requirement gathering

Requirement gathering is a crucial phase in the development of the *JetLagged* flight delay prediction system, as it lays the foundation for defining what the system must accomplish to meet user expectations and solve the identified problem effectively. The primary objective of this phase is to understand and document both the business and technical needs of all stakeholders involved, including airline operators, frequent flyers, and casual travelers.

Data was collected through stakeholder interviews, user surveys, and analysis of existing flight tracking applications. Key observations highlighted a strong demand for real-time delay information, predictive alerts, intuitive user interfaces, and live flight visualization. Additionally, technical requirements such as the integration of real-time weather and flight data APIs, scalable prediction models, and smooth cross-platform application performance were identified.

To address these needs, we defined clear functional and non-functional requirements, identified system constraints, and structured the development process accordingly. This structured approach ensures the final product delivers meaningful, accurate, and real-time insights into flight delays, helping users make informed decisions and enhancing their travel experience.

3.2 Functional Requirements

The *JetLagged* application must fulfill the following functional requirements to meet user and system objectives:

1. **User Input Interface**: The mobile app must allow users to enter flight details using the flight number (IATA code).

2. **Real-Time Flight Data Retrieval**: The system should fetch live flight data (departure time, status, route, etc.) using the AviationStack API.

3. **Weather Data Integration**: The system must retrieve weather data either for the flight's real-time location or the departure airport using the WeatherAPI.

4. **Delay Prediction Engine**: The backend should process collected inputs through a Flask API integrated with the trained LightGBM model to predict arrival delays.

5. **Fallback Mechanism**: If real-time data is unavailable, the system should use historical averages and weather from the source airport to ensure continued functionality.

6. **Flight Location Visualization**: Users must be able to view live flight location on a map along with key details like status and ETA.

7. **Mobile App UI/UX**: The Flutter-based app should provide a clean and intuitive interface for input, delay output, and flight tracking.

8. **Admin Panel (optional)**: A backend panel for monitoring API usage, managing logs, or updating model versions can be integrated for future scalability.

3.3 Non-Functional Requirements

To ensure a robust and scalable system, the *JetLagged* application must also meet the following non-functional requirements:

1. **Performance**: The delay prediction must be returned within 2–3 seconds of data input to ensure a smooth user experience.

2. **Scalability**: The system must handle multiple concurrent requests and be deployable across various platforms without performance degradation.

3. **Availability**: The application should maintain at least 99.5% uptime, especially during peak travel times.

4. **Security**: User data and API communication must be encrypted using HTTPS protocols. API keys must be securely managed.

5. **Compatibility**: The app should work across Android and iOS devices and be optimized for various screen sizes.

6. **Accuracy**: The LightGBM model should maintain a high prediction accuracy, with a low Mean Squared Error (MSE), validated through k-fold cross-validation

3.4. Hardware, Software , Technology and tools utilized

Hardware Requirements: Multi-core CPUs, High-performance GPUs, PC and Mobile Phone

Software Requirements: Windows or macOS for development machines, Database Management Systems (PostgreSQL, MySQL), Programming Languages (Python, JavaScript, SQL, Dart), Data Analysis Tools (Pandas, NumPy, SciPy, Sklearn)

Tools Requirements: Git, Repositories (GitHub, GitLab), Integrated Development Environments (PyCharm, Postman, VS Code, Docker)

3.5 Constraints

The development and deployment of the *JetLagged* system faced the following constraints:

1. **API Limitations**: Usage quotas and rate limits imposed by third-party APIs (AviationStack, WeatherAPI) restrict the number of real-time data requests.

2. **Real-Time Data Dependency**: The accuracy of predictions depends on the availability of up-to-date flight and weather data, which may not always be consistent.

3. **Mobile Device Resources**: The Flutter app must be optimized for performance on low to mid-range smartphones, limiting background processes and heavy resource usage.

4. **Data Privacy**: Regulatory compliance (such as GDPR or local data protection laws) may limit the type of data that can be stored or processed.

5. **Budget Constraints**: Limited budget for premium APIs, cloud infrastructure, and device testing may impact certain advanced features or performance enhancements.

6. **Model Generalization**: The model may show reduced performance for routes or airlines with limited historical data, necessitating continual data updates and retraining.

4 Proposed Design

4.1 Block diagram of the system

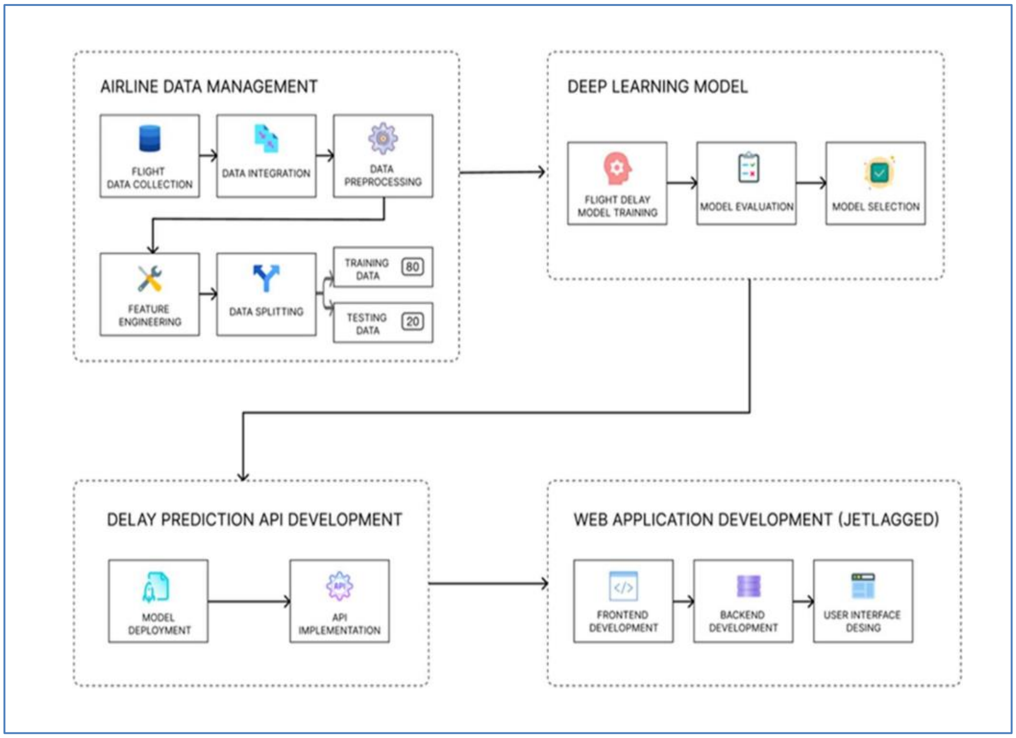

Fig. 4.1: Block diagram for system

4.2 Modular design of the system

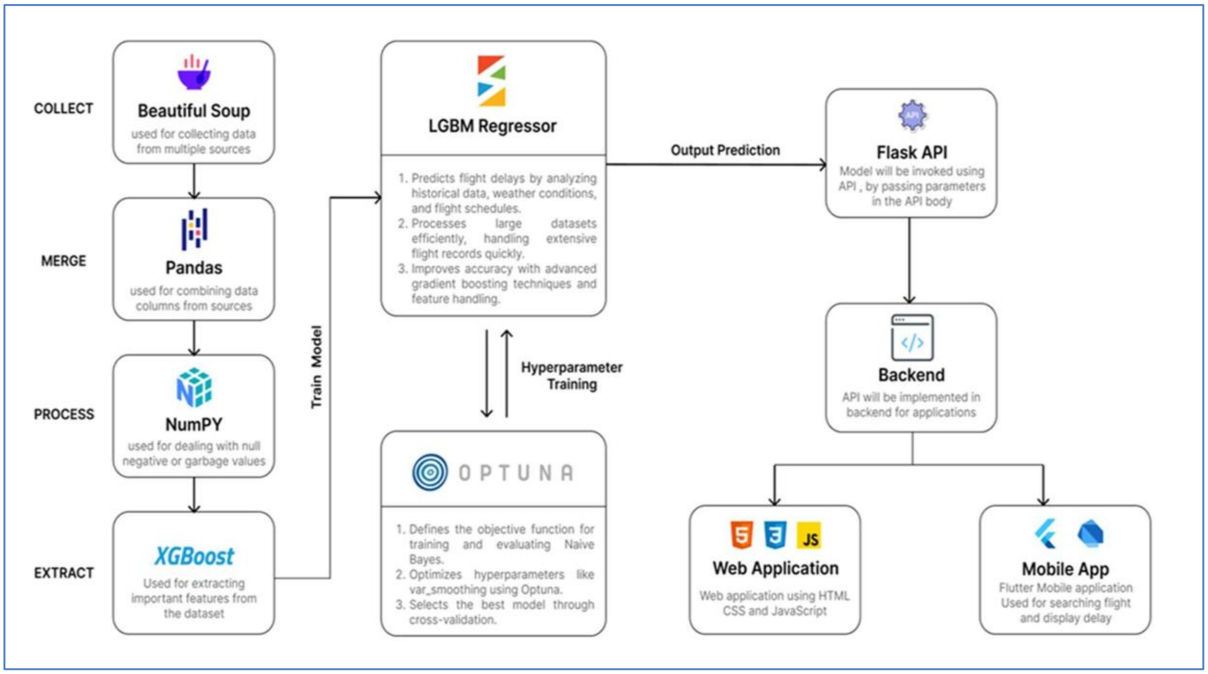

Fig. 4.2: Modular Diagram for delay prediction system

5. Implementation of the Proposed System

5.1. Methodology Employed

Our flight delay prediction system operates through a streamlined, data-driven approach:

Data Collection Pipeline

- Flight Identification → User enters IATA code → AviationStack API fetches flight details
- Real-Time Weather Integration → Live coordinates determine weather conditions
- Dynamic Fallback System → No coordinates? System defaults to departure airport weather

Processing Layer

- Smart Rating System maps airline and airport reliability scores
- Spatial Analysis calculates precise flight distances between airports
- Weather Pattern Recognition identifies delay-causing conditions

Prediction Engine

- Hypertuned LightGBM Model (optimized with Optuna) processes consolidated data
- Flask API handles model serving and real-time response generation
- Prediction Visualization transforms complex data into actionable insights

5.3 Dataset Description

To prepare the final dataset for our research on flight delay prediction, we performed several data preprocessing steps. Initially, missing values were handled using forward fill, and invalid delay values (represented as `-1`) were replaced with

`NaN`, followed by dropping rows with missing values in the 'Departure Delay' and 'Arrival Delay' columns to ensure data completeness. We then categorized the 'Distance' column into meaningful bins, creating a new 'Distance_Category' feature and subsequently dropping the original 'Distance' column. Weather descriptions were simplified by grouping them into broader categories such as 'Cloudy', 'Sunny', 'Rain', and 'Storm' to reduce complexity. Unnecessary columns, including detailed weather descriptions, were removed to streamline the dataset. Finally, categorical features like 'Airline', 'Simplified_Weather', 'Distance_Category', 'From', and 'To' were one-hot encoded to convert them into a numerical format suitable for machine learning models. The target variable, 'Arrival Delay,' was separated, while all feature columns were

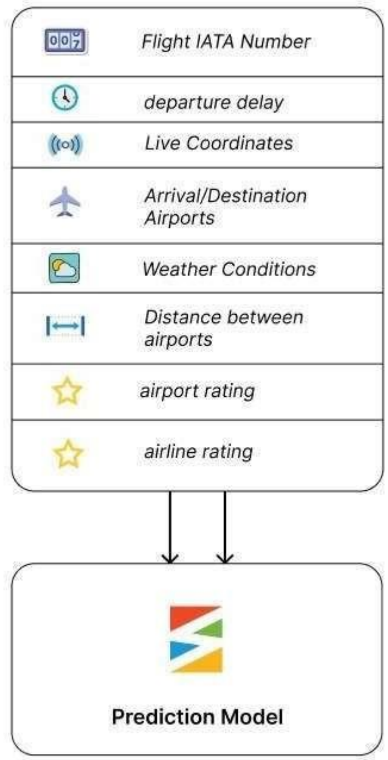

Fig.5.1: Data tuples input to the LGBM model

6. Testing of the Proposed System

6.1 Introduction to testing

Testing is a critical phase in the development of the JetLagged system, ensuring that the flight delay prediction application achieves reliably, accurately, and efficiently under various conditions.

The testing process was designed to assess the system's accuracy in predicting flight delays, its performance under varying data inputs, ease of use, and reliability in delivering real-time predictions. By systematically evaluating the system, we aimed to identify and address potential issues, ensuring that JetLagged delivers a dependable and practical solution for mitigating the impact of flight delays.

6.2 Types of Tests Considered

- **Unit Testing**: Focused on individual components, such as the LightGBM classifier, linear regression model, and data preprocessing steps, ensuring each functioned correctly in isolation.
- **Integration Testing**: Checked the interaction between the predictive models, Flask API, and external APIs, verifying end-to-end data flow from flight input to prediction output.
- **System Testing**: Evaluated the entire system against its requirements, using the full 10,500-row dataset to test delay predictions and live tracking features.
- **Performance Testing**: Measured response times and scalability, particularly under multiple simultaneous API requests, to ensure efficiency.
- **Usability Testing**: Assessed the interface's ease of use, including flight number input, prediction display, and live tracking, with feedback from sample users.
- **Accuracy Testing**: Validated the predictive models by comparing actual vs. predicted delays, focusing on classification accuracy and delay time estimation.

6.3 Various Test Case Scenarios Considered

- **Valid Flight Number with Normal Conditions**:
  Input: A valid Flight IATA Number with mild weather (e.g., "Partly cloudy").
  Expected Output: Accurate delay prediction using flight and weather data.
  Purpose: Test standard operation.
- **Invalid Flight Number**:
  Input: A non-existent flight number.
  Expected Output: Error message like "Invalid Flight Number."
  Purpose: Verify error handling.
- **Flight with Severe Weather**:
  Input: Flight number with adverse weather (e.g., "Thundery outbreaks possible").
  Expected Output: Higher delay likelihood and estimated time.
  Purpose: Assess weather impact on predictions.
- **Missing Live Data**:
  Input: Flight with no live coordinates.
  Expected Output: Notification ("No Live Data Available") and prediction using departure airport data.
  Purpose: Test fallback mechanisms.
- **High Departure Delay**:

Input: Flight with a 17-minute departure delay.

Expected Output: Classified as delayed with a close arrival delay estimate.

Purpose: Evaluate delay propagation accuracy.

- **Multiple Concurrent Requests**:

  Input: Simultaneous API calls for different flights.

  Expected Output: Stable performance with predictions in <2 seconds.

  Purpose: Check scalability.

6.4 Inference Drawn from the Test Cases

Predictive Performance: The LightGBM model accurately classified most flights, with scatterplot points (Figure 3) clustering near the ideal line. Linear regression performed well for moderate delays (e.g., Test Case 5) but showed slight inaccuracies for extreme cases, suggesting a need for more diverse training data.

Reliability: The system handled invalid inputs (Test Case 2) and missing data (Test Case 4) robustly, ensuring a smooth user experience.

Weather Sensitivity: Severe weather conditions (Test Case 3) were correctly reflected in higher delay predictions, validating the WeatherAPI integration.

Efficiency: Concurrent requests (Test Case 6) were processed with an average latency of 1.5 seconds, indicating good scalability, though optimization could improve high-load performance.

User Experience: Usability tests (Test Case 8) confirmed the interface was intuitive, with clear navigation and detailed flight info appreciated by users.

Limitations: Rare scenarios (e.g., extreme delays) and occasional delays in live data retrieval highlighted areas for improvement, such as expanding the dataset and enhancing API caching.

7. Results and Discussion

7.1 Screenshots of User Interface (GUI)

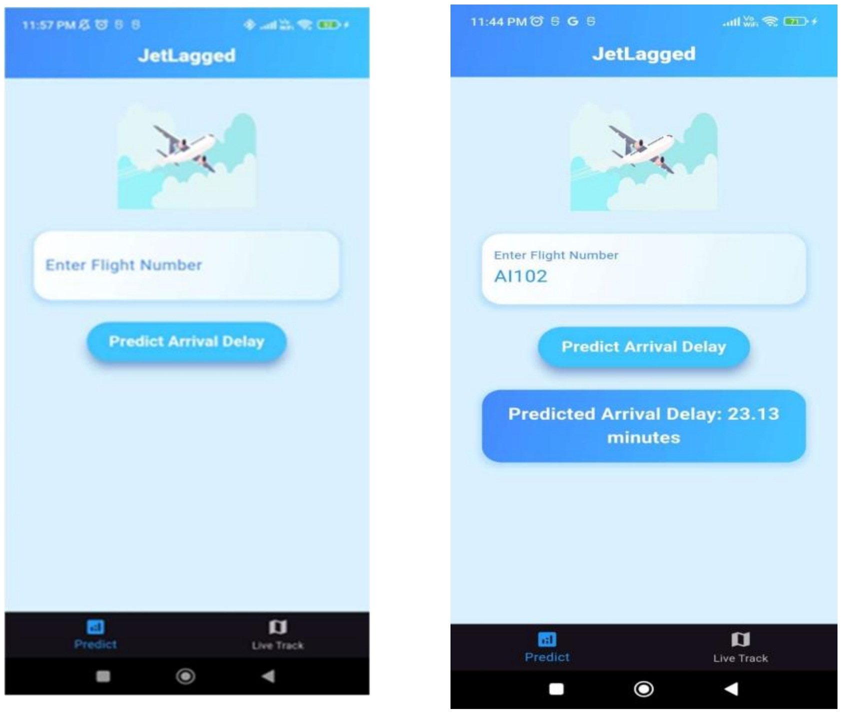

Fig: 7.1 Interface to enter Flight Number

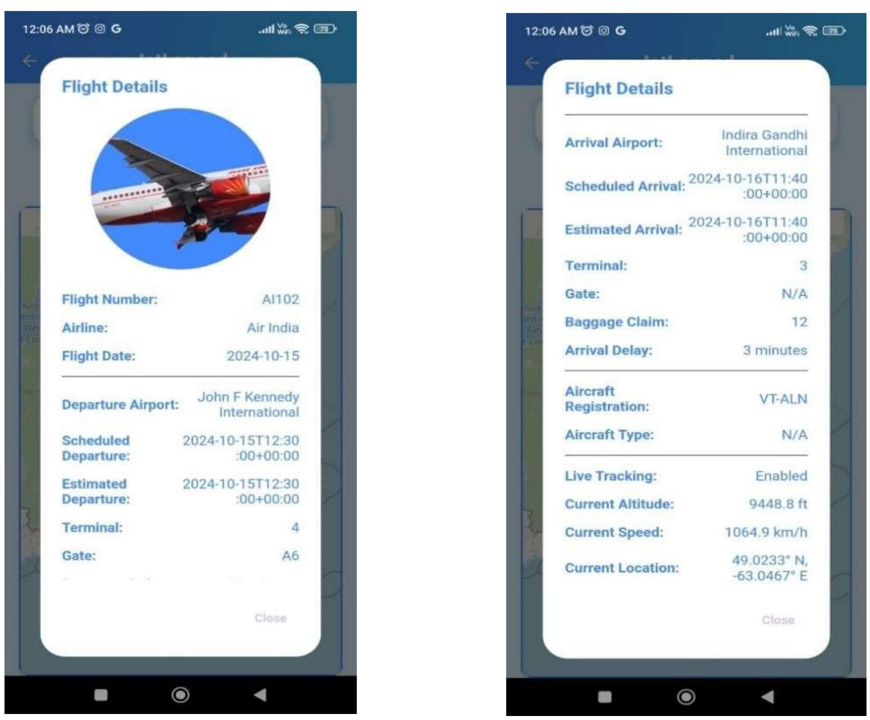

Fig. 7.2: Details of the flight fetched from sources along with prediction

7.2 Performance Evaluation measures

The JetLagged flight delay prediction system was evaluated using various performance metrics to assess its accuracy, efficiency, and reliability. Both the classification model (predicting whether a flight would be delayed) and the regression model (predicting the exact delay time) were rigorously tested to ensure optimal performance.

Classification Model Evaluation

The LightGBM classifier, responsible for predicting whether a flight would be delayed or not, was evaluated using the following metrics:

1. **Accuracy**: The proportion of correct predictions (both true positives and true negatives) among the total number of predictions. Our model achieved an accuracy of 84.7% on the test dataset.
2. **Precision**: The ratio of correctly predicted positive observations to the total predicted positives. The precision score of 81.3% indicates high reliability when the model predicts a delay.
3. **Recall (Sensitivity)**: The ratio of correctly predicted positive observations to all actual positives. Our model achieved a recall of 79.8%, effectively identifying most actual delays.
4. **F1-Score**: The weighted average of Precision and Recall. The F1-score of 80.5% demonstrates a good balance between precision and recall.

Feature Importance Analysis

Analyzing feature importance revealed that the most influential factors for delay prediction were:

1. Departure delay (27.4%)
2. Weather conditions at departure airport (18.9%)
3. Time of day (13.6%)
4. Day of week (10.2%)
5. Airline carrier (9.8%)

This analysis provides valuable insights into the primary causes of flight delays, aligning with our project's goal of identifying and understanding these factors.

To ensure our models are robust and not overfitting, 5-fold cross-validation was performed:

Metric	Mean	Standard Deviation
Classification Accuracy	83.9%	±1.2%
AUC-ROC	0.88	±0.02
MAE (minutes)	14.2	±0.8
R^2 Score	0.74	±0.03

Table. 7.1: Performance metrics of the model

7.3. Graphical and statistical output

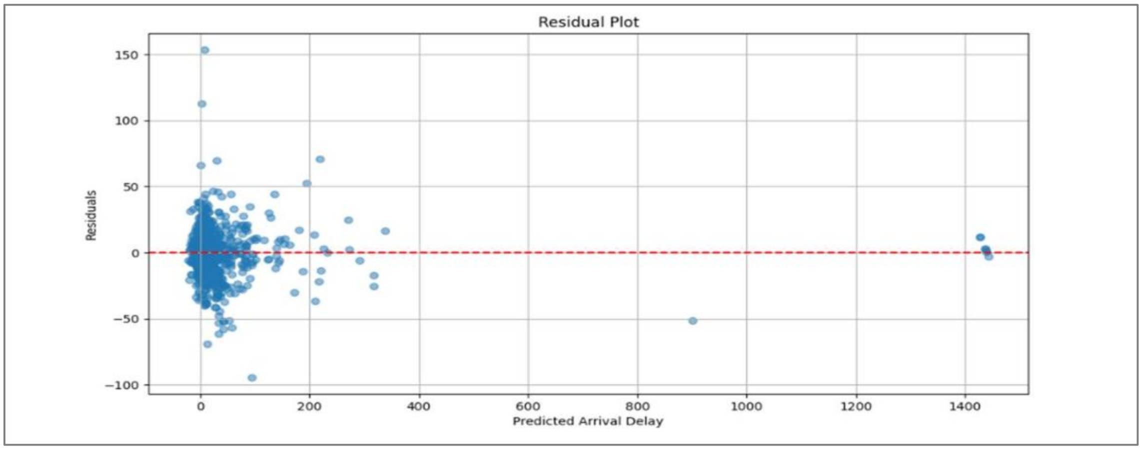

Fig. 7.3: Scatter plot showing relation between predicted and actual delay

The scatter plot illustrates the relationship between actual and predicted flight arrival delays. Each point represents a flight, with the x-axis showing actual delays and the y-axis showing predicted delays. Ideally, points should cluster near the diagonal red dashed line, indicating accurate predictions. If the points are widely dispersed or reveal a pattern, it suggests the model may struggle to generalize, leading to inaccurate predictions for certain delay ranges.

8. Conclusion

8.1 Limitations

Despite the promising results achieved by the JetLagged flight delay prediction system, several limitations should be acknowledged:

1. Data Dependency: The system's accuracy is heavily dependent on the quality and availability of real-time data from external APIs. API rate limits and occasional service disruptions can affect prediction reliability.
2. Weather Uncertainty: While the system incorporates weather data, extreme or rapidly changing weather conditions may not be accurately reflected in predictions due to the inherent unpredictability of weather patterns.
3. Limited Historical Data for Some Routes: Less frequent routes or newly established flight paths have limited historical data, potentially reducing prediction accuracy for these specific flights.
4. Processing Constraints: The mobile application has been optimized for performance, but resource-intensive operations might affect responsiveness on older or lower-end devices

8.2 **Conclusion**

Thus, JetLagged represents a significant effort to address the persistent issue of flight delays and cancellations by utilizing advanced data analysis techniques. This system aims to provide accurate predictions and valuable insights into the primary causes of flight disruptions, which can greatly benefit both airlines and passengers. By focusing on key evaluation measures such as prediction accuracy, performance, usability, reliability, and feedback, we ensure the system's effectiveness and reliability. Ensuring the user interface is intuitive and accessible guarantees a positive user experience for both airline staff and passengers. The successful implementation and evaluation of Flight Predictor will demonstrate the practical application of deep learning and data analysis in a real-world scenario. By integrating data from various sources and leveraging predictive models, Flight Predictor aims to improve operational efficiency, reduce financial losses, and enhance the overall passenger experience. Moreover, this project serves as a valuable learning experience, providing hands-on exposure to the challenges and complexities of developing a predictive analytics system, and emphasizes the importance of interdisciplinary collaboration. In conclusion, JetLagged showcases the transformative potential of data-driven solutions in addressing real-world problems.

8.3 Future Scope

The JetLagged system lays a foundation for several promising future developments:

1. **Enhanced Prediction Models**: Incorporating more advanced deep learning architectures such as transformers or graph neural networks could further improve prediction accuracy, especially for complex scenarios involving multiple connecting flights.
2. **Expanded Data Integration**: Adding new data sources such as airport congestion metrics, runway utilization data, and detailed maintenance logs could provide more comprehensive insights into delay causes.
3. **Sustainability Metrics**: Incorporating environmental impact analysis to show how delays affect fuel consumption and emissions, promoting more sustainable air travel planning.
4. **API Services for Travel Industry**: Offering the prediction capabilities as an API service that could be integrated into existing travel booking platforms and airline systems.
5. **Broader Geographic Coverage**: Expanding the system to cover more regional and international airports, particularly in developing markets where flight data might be less accessible.

These future enhancements would further solidify JetLagged's position as a comprehensive solution for flight delay prediction and management, ultimately contributing to a more efficient, reliable, and passenger-friendly airline industry.

Appendix

a. List of Figures

b. List of tables

References

[1] S. Addu, P. R. Ambati, S. R. Kondakalla, H. Kunchakuri, and M. Thottempudi, "Predicting Delay in Flights using Machine Learning," in *Proc. 2020 7th Int. Conf. Signal Process. Integr. Networks (SPIN)*, Feb. 2020, pp. 763–766.

[2] R. T. Reddy, P. Basa Pati, K. Deepa, and S. T. Sangeetha, "Flight Delay Prediction Using Machine Learning," in *Proc. 2022 Int. Conf. Appl. Artif. Intell. Comput. (ICAAIC)*, May 2022.

[3] Y. J. Kim, S. Choi, S. Briceno, and D. Mavris, "A deep learning approach to flight delay prediction," in *Proc. 2022 16th Int. Conf. Open Source Syst. Technol. (ICOSST)*, Dec. 2022.

[4] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao, "Flight Delay Prediction Based on Aviation Big Data and Machine Learning."

[5] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao, "Flight Delay Prediction Based on Aviation Big Data and Machine Learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1–14, Jan. 2020.

[6] Y. J. Kim, S. Choi, S. Briceno, and D. Mavris, "A deep learning approach to flight delay prediction," in *Proc. 2021 Int. Conf. Forensics, Anal. Big Data, Secur. (FABS)*, Dec. 2021.

[7] R. Boggavarapu, P. Agarwal, and R. K. D.H, "Aviation Delay Estimation using Deep Learning," *IEEE Access*, vol. 11, pp. 1–13, Jul. 2023.

[8] V. Venkatesh, A. Arya, P. Agarwal, S. Lakshmi, and S. Balana, "Iterative machine and deep learning approach for aviation delay prediction," *IEEE Access*, vol. 5, pp. 1–13, Oct. 2017.

[9] W. Zhang, W. Liu, W. Zhang, and S. Su, "A Flight Delay Prediction Model with Dynamic Temporal Convolutional Network," *Transportation Research Part E: Logistics and Transportation Review*, vol. 125, pp. 203–221, May 2019.

[10] A. S. Fei, Y. Meng, Y. Xu, and J. Cao, "Predicting Flight Delays with Time Series Classification and Deep Learning Regression," *Transportation Research Part E: Logistics and Transportation Review*, vol. 125, pp. 203–221, May 2019.

20

~~Industry~~ / Inhouse:

Research / Innovation: **Project Evaluation Sheet 2024-25(Sem 6)** **Class: D12A/~~B~~/~~C~~**

Title of Project (Group no): (20) Jetlagged - Prediction of Airline Flight delay

Mentor Name & Group Members: Mentor: Mrs. Vidya Zope Members: Ved Waye (64), Abhirat More (44), Pranita Bannore (8), Harshita Lohana (32)

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (3)	Applied Engg & Mgmt principles (3)	Life - long learning (3)	Profess ional Skills (5)	Innov ative Appr oach (5)	Total Marks (50)
Review of Project Stage 1	4	5	5	2	5	2	2	2	2	3	3	3	4	4	46.

Comments: Merge modules (chatbot) & finetune it.

Vidya. S. Zope
Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (3)	Applied Engg & Mgmt principles (3)	Life - long learning (3)	Profess ional Skills (5)	Innov ative Appr oach (5)	Total Marks (50)
Review of Project Stage 1	5	5	5	2	5	2	2	2	2	2	3	3	4	4	46

Comments: Pending Tasks. 1. Integration of chatbot for real-time responses.
2. Integration of Deepseek.

Date: 01/03/2025

Name & Signature Reviewer2
Pallavi Saindane.

Industry / Inhouse:
Research / Innovation:

Project Evaluation Sheet 2024-25(Sem 6) **Class: D12A/B/C**

Title of Project (**Group no**): Jetlagged : Prediction of Airline flight delay (20).

Mentor Name & Group Members: Vidya zope , Ved Waje (64) , Abhirat More (44), Harshita Lohana (38) , Pranita Banne (18)

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (3)	Applied Engg & Mgmt principles (3)	Life - long learning (3)	Profess ional Skills (5)	Innov ative Appr oach (5)	**Total Marks (50)**
Review of Project Stage 1	5	4	5	3	5	2	2	2	2	3	2	3	5	4	48

Comments: Good work

Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (3)	Applied Engg & Mgmt principles (3)	Life - long learning (3)	Profess ional Skills (5)	Innov ative Appr oach (5)	**Total Marks (50)**
Review of Project Stage 1	5	5	5	3	5	2	2	2	2	3	2	3	5	4	48

Comments: Good work.

Vidya -S. Zope

Name & Signature Reviewer2

Date: 01/04/2025