**0th Review Report – Flight Delay Prediction**

**1. Problem Identification**

Flight delays have significant negative impacts on airline operations, airport efficiency, and passenger satisfaction. These delays can be caused by various factors such as weather conditions, air traffic congestion, mechanical issues, and operational inefficiencies. Traditional delay prediction models often fail to capture the complex relationships among these factors, leading to unreliable predictions.

This project aims to leverage machine learning techniques, specifically Gradient Boosted Decision Trees (XGBoost), to predict flight delays with higher accuracy. By transitioning from a binary classification model to a regression-based approach, we aim to provide precise delay duration estimates rather than a simple delayed/not delayed outcome.

**2. Objective and Scope of the Project Work**

**Objective:**

The primary objective of this project is to develop a machine learning-based prediction model that accurately forecasts flight delays in minutes. The model will assist airlines, airports, and passengers in better managing schedules and mitigating disruptions.

**Scope:**

* **Data Collection & Preprocessing:** Gathering historical flight data, weather reports, and airport congestion information.
* **Feature Engineering:** Extracting relevant factors such as time of day, airline performance, and weather conditions.
* **Model Development:** Implementing XGBoost for flight delay regression modeling.
* **Performance Evaluation:** Using metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) to assess prediction accuracy.
* **Implementation & Deployment:** Potential deployment as a web or mobile-based decision support tool for airlines.

**3. Literature Relevancy to the Topic**

Several research studies have explored machine learning approaches for flight delay prediction:

1. **Flight Delay Prediction Using Machine Learning** – R.T. Reddy et al. applied machine learning techniques to predict flight delays at JFK airport, focusing on weather-related factors.
2. **Predicting Delay in Flights using Machine Learning** – This study compared multiple algorithms, including Decision Trees, Random Forest, and Neural Networks, for delay prediction.
3. **Flight Delay Predictions and the Study of Its Causal Factors** – Investigated both the prediction and the contributing factors influencing delays.
4. **Enhancing Airline Operations by Flight Delay Prediction** – Focused on optimizing airline scheduling based on machine learning predictions.
5. **Flight Delay Prediction Based on Gradient Boosting Ensemble Models** – Demonstrated the effectiveness of XGBoost in predicting flight delays with improved accuracy.

**4. Literature Gap Analysis**

While existing research has significantly contributed to flight delay prediction, gaps remain:

* **Limited focus on regression models:** Many studies treat delay prediction as a classification problem (delayed/not delayed) rather than estimating actual delay time.
* **Insufficient feature integration:** Most models fail to incorporate real-time weather conditions, air traffic congestion, and operational constraints.
* **Model Generalization:** Some studies achieve high accuracy on specific datasets but struggle to generalize across different regions and airports.
* **Deployment Challenges:** Few studies focus on the practical implementation of these models for real-time use by airlines and airports.

**Proposed Solution**

To address these gaps, this project:

1. Uses a **regression-based approach (XGBoost)** to predict exact delay times.
2. Integrates **multiple data sources** such as historical flight records, weather data, and airport congestion levels.
3. Implements **advanced feature engineering techniques** to improve model accuracy.
4. Focuses on **scalability and real-world applicability** for potential deployment as a decision-support tool.

This concludes the **0th Review Report**. Once approved, I will begin working on the **1st Review Report** covering architecture, algorithms, modules, and results.