✈️ Airlines Flight Delay Prediction – Project Overview

---

1. Abstract

Flight delays are a common issue in the aviation industry, resulting in significant economic losses and passenger dissatisfaction.
This project applies **machine learning techniques** on airline flight Data to **predict flight delays** based on factors such as airline, route, departure time, weather, and distance.
By analyzing patterns and building predictive models, the project aims to assist airlines in **better scheduling and resource optimization**.

---

2. Introduction

- Flight delays are influenced by multiple factors (air traffic, weather, operational inefficiencies).
- Predicting delays helps minimize disruptions.
- The project focuses on **binary classification**:
  - **0 = On-time**
  - **1 = Delayed**

---

3. Dataset Description

- **Source**: Airlines Flight Dataset (CSV/Excel).
- **Size**: ~X rows, Y features (replace with actual).
- **Features** (examples):
  - Airline, Flight Number, Departure Time, Arrival Time, Distance, Origin Airport, Destination Airport, Weather Condition.
- **Target Variable**: Delay (0/1).

---

4. Workflow Diagram

(Insert a simple flowchart – I can generate one for you if you like)

Workflow:
Data Collection → Data Preprocessing → Exploratory Data Analysis → Feature Engineering → Train/Test Split → Model Building → Model Evaluation → Deployment

---

5. Methodology

5.1 Data Preprocessing

- Handle missing values (mean/median).
- Remove duplicates.

- Encode categorical variables (Airline, Airport).
- Scale continuous features (StandardScaler / MinMaxScaler).

5.2 Exploratory Data Analysis (EDA)

- Distribution of delays across airlines & airports.
- Impact of flight distance and departure time.
- Correlation heatmap to detect feature relationships.

5.3 Feature Engineering

- Create derived features:
    - **Peak Hour Indicator**
    - **Flight Duration Category**
    - **Weather Impact Score**

---

6. Models Used

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- XGBoost

---

7. Model Evaluation

- Metrics used: **Accuracy, Precision, Recall, F1-score, ROC-AUC**.
- Confusion Matrix for classification errors.
- Cross-validation for robust evaluation.

---

8. Results & Key Findings

- Random Forest and XGBoost showed the **best performance** (higher accuracy & F1).
- Delay probability is strongly affected by:
    - Departure time (late-night/early-morning flights are less delayed).
    - Weather conditions.
    - Busy airports and long routes.

---

9. Key Takeaways

- Machine learning models can effectively predict delays.

- Weather and airport congestion are critical features.
- Airlines can use such models for **better planning & passenger satisfaction**.
- Future improvements: Incorporate **real-time weather & traffic data**.

---

10. Conclusion

This project demonstrates the use of machine learning in the aviation sector for predicting delays.
The models provide valuable insights that can support **decision-making, scheduling, and customer service improvements**.