Synopsis on Flight Delay Prediction

Team Members –

1. Dewang Bedarkar - 240844525007
2. Gaurav Gulia - 240844525009
3. Hitesh Thombare - 240844525017
4. Narendra Gadhe - 240844525020
5. Omkar Burute - 240844525023
6. Pranay Shah - 240844525026
7. Saurabh Patil - 240844525036
8. Vaishnavi Bhure - 240844525044

Github Repo Link – [Flight_Delay_Predictor_Team1](#)

Problem Statement:

This project develops a machine learning model to predict flight delays using historical data, including airline, flight schedules, and airport congestion. The outcome will include a predictive model, a UI for flight delay insights, and a dashboard for visualizing historical data trends.

Deliverables :

Trained Model:

- Optimized machine learning model for predicting flight delays (saved as .pkl).

Feature Importance Analysis:

- Key insights with visualizations (e.g. bar charts, pairplots, heatmaps) showing influential factors.

Evaluation Metrics:

- Performance report with metrics like accuracy, precision, recall (classification) or MAE, RMSE (regression).

Prediction Pipeline:

- End-to-end pipeline for data preprocessing and generating predictions.

Prototype UI:

- A simple interface for real-time flight delay predictions (using Streamlit or Flask).

Code Documentation:

- Detailed steps to train, evaluate, and integrate the model.

Data Dictionary:

Column Name	Description
Year	Year of the flight.
Quarter	Quarter of the year (1-4).
Month	Month of the flight.
DayofMonth	Day of the month.
DayOfWeek	Day of the week.
FlightDate	Flight date in the format yyyymmdd.
Marketing_Airline_Network	Unique marketing carrier code. May include numeric suffixes (e.g., PA, PA(1), PA(2)) to distinguish different carriers. Use for cross-year analysis.

Operated_or_Branded_Code_Share_Partners	Reporting carrier operated or branded code share partners.
DOT_ID_Marketing_Airline	Identification number assigned by US DOT for a unique airline (carrier).
IATA_Code_Marketing_Airline	Code assigned by IATA for identifying a carrier. The same code may be reused by different carriers over time.
Flight_Number_Marketing_Airline	Flight number for the marketing airline.
Originally_Scheduled_Code_Share_Airline	Unique scheduled operating carrier code with possible numeric suffixes.
DOT_ID_Originally_Scheduled_Code_Share_Airline	DOT identification number for the originally scheduled code-share airline.
IATA_Code_Originally_Scheduled_Code_Share_Airline	IATA code for the originally scheduled code-share airline.
Flight_Num_Originally_Scheduled_Code_Share_Airline	Flight number for the originally scheduled code-share airline.
Operating_Airline	Unique carrier code for the operating airline, with possible numeric suffixes.
DOT_ID_Operating_Airline	DOT identification number for the operating airline.
IATA_Code_Operating_Airline	IATA code for the operating airline.
Tail_Number	Aircraft tail number.
Flight_Number_Operating_Airline	Flight number for the operating airline.
OriginAirportID	Unique ID for the origin airport, assigned by US DOT.
OriginAirportSeqID	Sequence ID for the origin airport, identifying time-specific details.
OriginCityMarketID	City market ID for the origin airport, consolidating airports serving the same city market.
Origin	Origin airport code.
OriginCityName	Name of the city for the origin airport.
OriginState	State code for the origin airport.
OriginStateFips	State FIPS code for the origin airport.
OriginStateName	Full state name for the origin airport.
OriginWac	World area code (WAC) for the origin airport.
DestAirportID	Unique ID for the destination airport, assigned by US DOT.
DestAirportSeqID	Sequence ID for the destination airport, identifying time-specific details.
DestCityMarketID	City market ID for the destination airport, consolidating airports serving the same city market.
Dest	Destination airport code.
DestCityName	Name of the city for the destination airport.
DestState	State code for the destination airport.
DestStateFips	State FIPS code for the destination airport.
DestStateName	Full state name for the destination airport.
DestWac	World area code (WAC) for the destination airport.
CRSDepTime	Scheduled departure time (local time, hhmm).
DepTime	Actual departure time (local time, hhmm).
DepDelay	Difference in minutes between scheduled and actual departure time. Negative values indicate early departures.
DepDelayMinutes	Departure delay in minutes. Early departures are set to 0.
DepDel15	Departure delay indicator (1 = delay of 15 minutes or more).
DepartureDelayGroups	Departure delay intervals (in 15-minute increments up to 180).
DepTimeBlk	Scheduled departure time block (hourly intervals).
TaxiOut	Taxi-out time in minutes.
WheelsOff	Wheels-off time (local time, hhmm).
WheelsOn	Wheels-on time (local time, hhmm).
TaxiIn	Taxi-in time in minutes.

CRSArrTime	Scheduled arrival time (local time, hhmm).
ArrTime	Actual arrival time (local time, hhmm).
ArrDelay	Difference in minutes between scheduled and actual arrival time. Negative values indicate early arrivals.
ArrDelayMinutes	Arrival delay in minutes. Early arrivals are set to 0.
ArrDel15	Arrival delay indicator (1 = delay of 15 minutes or more).
ArrivalDelayGroups	Arrival delay intervals (in 15-minute increments up to 180).
ArrTimeBlk	Scheduled arrival time block (hourly intervals).
Cancelled	Flight cancellation indicator (1 = Yes).
CancellationCode	Reason for flight cancellation.
Diverted	Diverted flight indicator (1 = Yes).
CRSElapsedTime	Scheduled elapsed flight time in minutes.
ActualElapsedTime	Actual elapsed flight time in minutes.
AirTime	Actual flight time in minutes.
Flights	Number of flights.
Distance	Distance between origin and destination airports (in miles).
DistanceGroup	Distance intervals (every 250 miles).
CarrierDelay	Delay due to carrier issues (in minutes).
WeatherDelay	Delay due to weather issues (in minutes).
NASDelay	Delay due to National Air System issues (in minutes).
SecurityDelay	Delay due to security issues (in minutes).
LateAircraftDelay	Delay due to late-arriving aircraft (in minutes).

Architecture Diagram: