

REPORT

Shipment Delay Prediction

1. Project Overview

The aim of this project is to predict the likelihood of a shipment being delayed based on factors such as delay duration and weather conditions. The dataset consists of features like Shipment ID, Origin and Destination cities, Shipment Date, Planned and Actual Delivery dates, Vehicle Type, Distance, Weather and Traffic conditions, and Delay (the target variable).

The project's primary goal is to predict delays and understand the impact of external factors, particularly weather, on delivery times.

2. Dataset

The dataset contains the following features:

- **Shipment ID:** Unique identifier for each shipment.
- **Origin:** Origin city of the shipment.
- **Destination:** Destination city of the shipment.
- **Shipment Date:** Date on which the shipment was made.
- **Planned Delivery:** Expected delivery date.
- **Actual Delivery:** Actual date of delivery.
- **Vehicle Type:** Type of vehicle used for transportation (Truck, Lorry, Container, Trailer).
- **Distance (km):** Distance between origin and destination.
- **Weather Conditions:** Weather status during transportation (Clear, Rain, Fog, etc.).
- **Traffic Conditions:** Traffic conditions during the shipment (Light, Moderate, Heavy).
- **Delay:** Target variable indicating whether the shipment was delayed (Yes/No).
- **Delay Duration:** Duration of delay (in hours/days). *#added using formula*

3. Methodology

3.1. Data Preprocessing

Data was preprocessed by:

- Handling missing values and categorical variables.
- Encoding categorical variables such as Weather Conditions and Traffic Conditions.
- Feature scaling was performed on numerical data like Distance and Delay Duration.

3.2. Model Development

We used a machine learning model to predict the likelihood of delays. The features used for prediction were:

- **Delay Duration:** The time by which the shipment was delayed.
- **Weather Conditions:** The weather impact on delivery performance.

We chose classification algorithms like Logistic Regression, Random Forest, and SVC (Support Vector Classifier) to predict the delay.

3.3. Evaluation

Models were evaluated using metrics such as accuracy, precision, recall, and F1-score to determine the best performing model for predicting delayed shipments.

3.4. Flask Integration

For ease of use and deployment, the model was integrated with Flask in a Jupyter Notebook environment. A simple web application was created where users could input shipment details (e.g., Delay Duration and Weather Conditions) and get a prediction on whether the shipment will be delayed.

- The Flask app was set up to accept input from the user in a form format.
- The input was processed, and the prediction was generated based on the trained model.
- The model prediction was displayed on the web page as "Delayed" or "On Time".

4. Results

The best performing model was selected based on evaluation metrics. The Flask application successfully predicted whether shipments would be delayed based on input values for delay duration and weather conditions.

5. Conclusion

This project demonstrates how external factors like weather conditions can impact shipment delays. By using delay duration and weather conditions, we were able to accurately predict whether a shipment would be delayed. The integration of the model into a Flask application made it easy to deploy and use for real-time predictions.

Future work may involve incorporating other features, such as traffic conditions and vehicle types, into the model for more accurate predictions. Additionally, integrating the system with a database to handle larger datasets and make predictions at scale could further enhance the project.