REPORT — Amazon Delivery Time Estimation System

1. Project Title

Amazon Delivery Time Estimation System using Machine Learning and Streamlit

---

2. Aim

The aim of this project is to design and implement a predictive system that estimates delivery times for Amazon orders using real-world delivery data.
It combines machine learning for prediction and Streamlit for visualization, allowing users to analyze, predict, and monitor delivery performance effectively.

---

3. Technologies Utilized

Languages & Tools: Python, Streamlit, SQLite3

Development Platforms: Google Colab (model building), VS Code (dashboard creation)

Libraries: Pandas, NumPy, Plotly, Joblib, Scikit-learn

Visualization: Plotly Express interactive charts

Storage: SQLite database for saving predicted results

---

4. Dataset Overview

The project uses an Amazon delivery dataset that contains:

Store and Drop coordinates

Agent age and ratings

Traffic and weather conditions

Product category and delivery area

Target variable: Delivery_Time

Using the Haversine distance formula, the project calculates distance_km between store and delivery point to make predictions more accurate.

---

5. Implementation Steps

Data Preparation

Imported the dataset and handled null/missing values.

Generated distance feature using latitude and longitude.

Encoded categorical columns (Weather, Traffic, Vehicle, etc.) into machine-readable format.

Model Building

Several regression algorithms were evaluated.

The best model (based on accuracy and error metrics) was finalized and exported as a .pkl file.

Model evaluation metrics were analyzed on test data to confirm reliability.

Interface Development

Built an interactive Streamlit dashboard (app2.py).

Integrated five modules/pages:

Overview: Explore delivery data visually.

Prediction: Input features to predict delivery time.

Search & Filters: Retrieve and filter previous predictions.

Analytics: Visualize insights through plots.

About: Summarize the project.

Database Management

Implemented SQLite to store every prediction instance with delivery conditions.

Enabled filtering by weather, traffic, or area for analytical queries.

---

6. Visualizations

Histogram: Delivery time distribution.

Boxplots: Delivery time across traffic and weather conditions.

Scatter plot: Distance vs delivery time relation.

Correlation heatmap: Numeric relationships between variables.
These visuals help in understanding the impact of different factors on delivery duration.

---

7. Major Highlights

Interactive multi-page Streamlit app with animated gradient background.

Real-time prediction using pre-trained machine learning model.

Database-backed system for saving results and performing advanced analytics.

Attractive, user-friendly dashboard UI.

Lightweight and fully executable on local machine.

---

8. Benefits

Supports logistics management with predictive insights.

Enables delivery agents and planners to forecast expected delays.

Improves operational decision-making for large-scale delivery networks.

Demonstrates real-world integration of AI and analytics in e-commerce.

---

9. Outcome

The project successfully predicts delivery time based on several contextual and environmental parameters.
The dashboard acts as both a decision-support tool and a data visualization platform, showcasing how AI can optimize e-commerce logistics.

---

10. Developed By

Name: MRITYUNJAY ACHARYA
Domain: Artificial Intelligence & Machine Learning
Tools Used: Streamlit, Python, Scikit-learn, SQLite, Plotly
Project Type: End-to-End Predictive Analytics Dashboard