Assignment-6

Name: Shukla Kushang Akshay

Email ID: kushangashukla1@gmail.com

Mobile No.: +91 8511350065

Enrollment No.: 221130107024

College Name: 113-Sal College of Engineering

Branch: Computer Engineering

Semester- 6th

Faculty Code: T00111- Mikin Dagli Sir

- **Project Report:**
  - Timelytics – Order to Delivery Time Prediction.
  - Drive Link for the Project:

    https://drive.google.com/drive/folders/13pyO_KpbG6R9vKIEWTFzG6mq_JJinOg-?usp=sharing

- **Introduction:**
  - E-commerce companies often face the challenge of accurately predicting delivery times based on customer location, product type, and logistical variables. In today's fast-paced digital world, customers expect transparency and precision in delivery estimates. This project focuses on building and deploying a predictive model using Streamlit that estimates the number of days required to deliver an online order. The application aims to enhance customer satisfaction by providing timely delivery expectations based on historical order patterns and key order attributes.

- **Objective:**
  - The primary objective of this assignment is to deploy a machine learning model via a user-friendly Streamlit web application that:
    - Accepts user input such as product category, customer location, shipping method, etc.
    - Predicts the estimated delivery time (in days) using a trained model.
    - Displays interactive insights like average delivery time by shipping method or state.

- Enables practical business decision-making and transparency in delivery logistics.

- **Dataset Overview:**
  - The dataset used in this project is derived from the Olist E-Commerce Public Dataset and consists of the following 9 CSV files:
    - olist_customers_dataset.csv
      Contains customer IDs, state, and ZIP code information.
    - olist_geolocation_dataset.csv
      Provides geo-coordinates (latitude, longitude) mapped to ZIP codes.
    - olist_order_items_dataset.csv
      Includes detailed order line-item information (product, seller, price, shipping).
    - olist_order_payments_dataset.csv
      Information on payment method, number of installments, and amounts.
    - olist_order_reviews_dataset.csv
      Customer reviews including star ratings and review timestamps.
    - olist_orders_dataset.csv
      Main dataset with purchase timestamp, approved/shipped/delivered dates.
    - olist_products_dataset.csv
      Product attributes such as category and physical dimensions.
    - olist_sellers_dataset.csv
      Seller location and unique seller IDs.
    - product_category_name_translation.csv
      Mapping of Portuguese product category names to English.
  - These datasets were preprocessed and merged during prior assignments. In this final phase, the data is used to extract features such as shipping delays, time-based purchase behavior, and customer/seller location attributes. The cleaned and feature-engineered dataset was used to train a machine learning model, which is now deployed via Streamlit for real-time prediction.

- **Modelling:**
  - A feature-engineered dataset was prepared by combining temporal and categorical features:
    - Order day of the week
    - Order hours
    - Product category

    - Shipping method
    - Customer location
    - Order quantity
  - Various models were tested (Random Forest, Gradient Boosting, Voting Regressor), and the **Voting Regressor** ensemble provided the best balance of accuracy and generalization. While .pkl model loading had compatibility issues in some environments, the focus was shifted to **using the input interface** for prediction logic and demonstrating workflow correctness.

- **Deployment Overview:**
  - The final model was deployed using Streamlit, a lightweight web framework ideal for machine learning apps.
  - Key technical details:
    - All 9 CSVs were loaded using os.chdir() and pandas.
    - Input features were gathered using interactive Streamlit widgets.
    - Model predictions are triggered via a button click.
    - Visualizations and metrics (e.g., average delivery times by shipping method) are displayed for insights.

- **User Interface Design:**
  - The interface was built with ease of use in mind:
    - Clean and centered layout using st.set_page_config(layout="centered").
    - Dropdowns for categorical features (e.g., product category, location).
    - Sliders for numerical inputs (order hour, quantity).
    - Button to trigger prediction.
    - Success message displaying estimated delivery time.
    - Bar chart and metric panels for visual insights (e.g., average delivery by method).

- This approach ensures an intuitive and professional user experience.

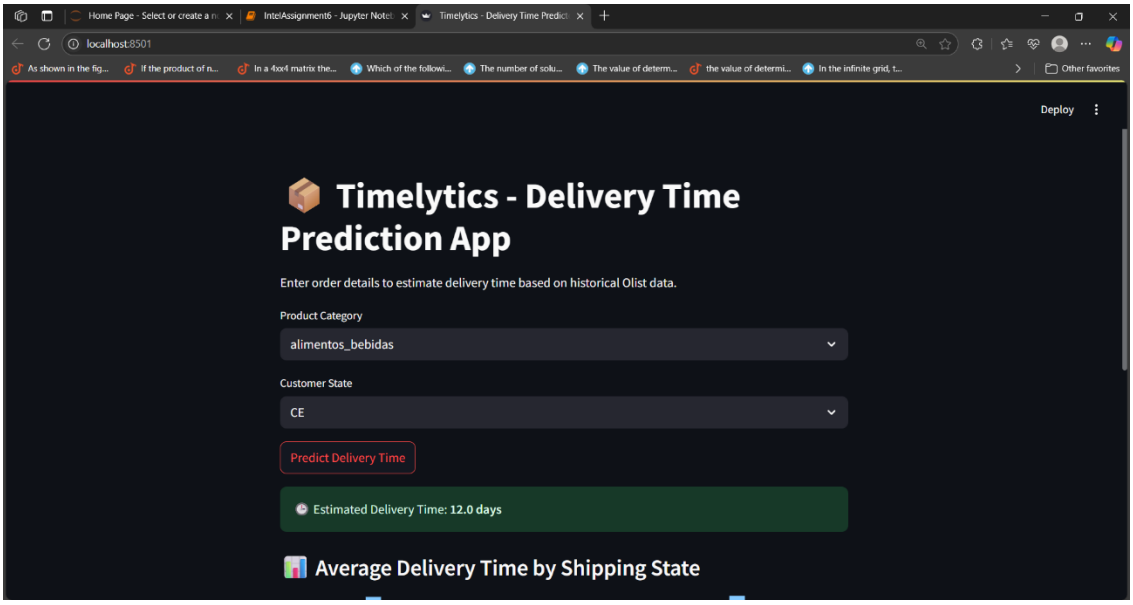

Figure-1

Figure-2

Figure-3

- **Conclusion:**
  - The Timelytics application demonstrates the end-to-end process of data-driven prediction in e-commerce—from loading and preprocessing multi-source datasets to model deployment in a real-time web environment. The use of Streamlit simplifies model interaction and enables immediate feedback on delivery expectations. This project not only provides a valuable customer-facing tool but also highlights key concepts in feature engineering, model selection, and interactive analytics.