ACROPOLIS INSTITUTE OF TECHNOLOGY AND RESEARCH

Department of Computer Science & Engineering (Data Science)

Synopsis

On

E-Commerce Sales Prediction

1. INTRODUCTION

1.1 Overview

The E-commerce Sales Analysis and Prediction System is a data-driven software solution designed to analyze online shopping data, visualize trends, and build machine learning models to predict future sales, revenue, customer behavior, and product demand.

This project helps in understanding customer buying patterns, popular product categories, seasonal sales trends, and predicting future demand using historical data.

1.2 Purpose

The purpose of this project is to analyze real-world e-commerce sales data to uncover meaningful insights and use machine learning techniques to predict future outcomes such as sales trends, revenue, customer behavior, and delivery performance. In today's rapidly evolving digital marketplace, e-commerce platforms generate vast amounts of data from customer purchases, payments, product categories, and logistics. However, this raw data is rarely used to its full potential. Through this project, the aim is to transform raw data into valuable business intelligence by applying data preprocessing, exploratory data analysis (EDA), visualization, and predictive modeling. This system will help businesses make informed decisions about inventory management, marketing strategies, pricing models, and customer satisfaction. Ultimately, the project intends to build a smart, data-driven solution that not only visualizes trends but also forecasts future sales and enhances the overall efficiency of e-commerce operations.

2. LITERATURE SURVEY

2.1 Existing Problem

Although e-commerce platforms collect large volumes of data, most businesses do not fully utilize it for analytical or predictive purposes. Traditional sales tracking systems only focus on recording transactions and generating static reports, without offering deeper insights into customer preferences, product demand, seasonal trends, or future revenue predictions. Many small and medium-sized businesses still rely on manual spreadsheets, which are time-consuming, error-prone, and provide no predictive intelligence. There is no automated system to analyze customer buying behavior, predict sales, detect delivery delays, or recommend inventory stocking levels. Additionally, without proper visualization tools, identifying important trends such as the most profitable products, high-return items, or cities with maximum sales becomes difficult. As a result, companies often make business decisions based on assumptions rather than data-driven insights, leading to losses, poor customer satisfaction, and inefficient use of resources.

2.2 Proposed Solution

To overcome these challenges, this project proposes a complete data-driven E-commerce Sales Analysis and Prediction System using Python, data analytics, and machine learning. The system will collect historical e-commerce data, either from real-world sources like the Olist dataset (Kaggle) or through synthetic datasets. The data will go through preprocessing steps such as cleaning, handling missing values, formatting dates, and generating new useful features like monthly sales, revenue per customer, discount impact, and delivery time. Exploratory Data Analysis (EDA) will be performed to identify patterns in sales, customer demographics, product performance, and revenue distribution. Interactive visualizations will be created using Matplotlib or Plotly to display trends clearly. For the predictive part, machine learning models such as Linear Regression, Random Forest will be used to predict sales or order amounts based on past data. The system can also predict late deliveries, customer churn, or return probability. This solution will help businesses make smarter, data-backed decisions and gain a competitive advantage in the e-commerce industry.

3. THEORETICAL ANALYSIS

3.1 Block Diagram
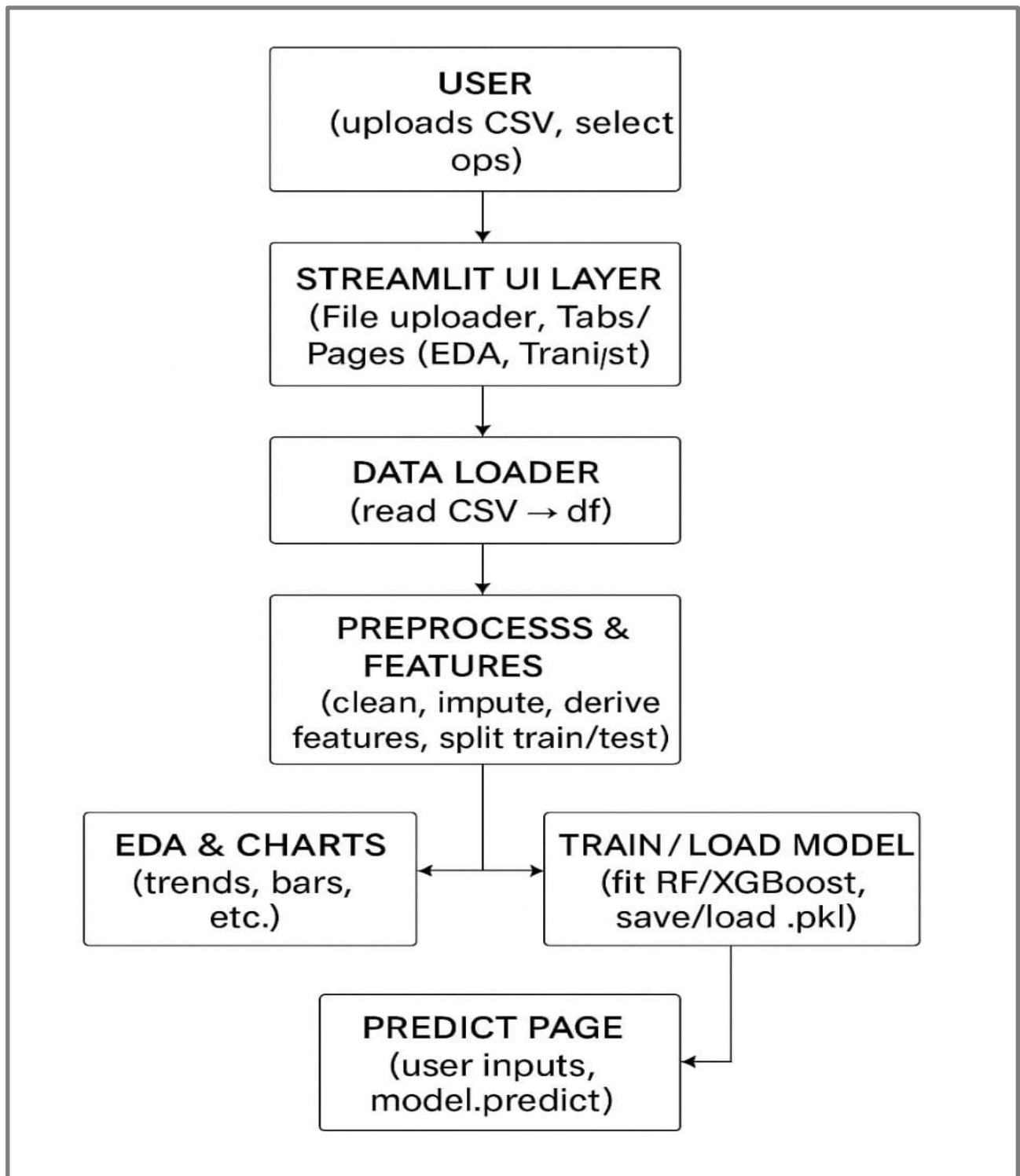
Fig.1. Workflow of E-Commerce Sales Analysis

3.2 Hardware/Software Designing

Hardware Requirements:

- Processor: Intel i3/i5/i7 or Ryzen 3/5

- RAM: 8GB

- Storage: 128 GB HDD or more

- Network: 10/100 Mbps Internet Connection

- OS: Linux-Based or Windows Server

Software Requirements:

- Programming Language: Python

- Libraries: Pandas, NumPy, Matplotlib, Seaborn

- ML frame works: Scikit-Learn / XGBoost

- Visualization: Matplotlib / Plotly / Power BI / Streamlit

- IDE / Tools: Jupyter Notebook / VS code

- Dataset: E-commerce Dataset (Kaggle)

- Version Control: Git with GitHub

4. APPLICATIONS

- **Sales Forecasting and Revenue Prediction**
  Businesses can predict future sales and revenue using historical data trends. This helps in planning budgets, marketing strategies, and business expansion.

- **Inventory and Stock Management**
  By predicting product demand, businesses can maintain the right amount of stock, preventing overstocking or stockouts, reducing storage costs, and improving supply-chain efficiency.

- **Customer Behavior and Personalization**
  The system helps in analyzing customer preferences, buying frequency, and spending patterns. This allows companies to personalize recommendations and improve customer experience.

- **Delivery and Logistics Performance Monitoring**
  Delivery time and performance can be tracked and predicted. This helps identify delays, improve courier selection, and enhance customer satisfaction.

- **Pricing and Discount Strategy Optimization**
  By studying the relationship between discounts and sales, businesses can determine the best discount rates that increase sales without reducing profit margins.

- **Return and Refund Analysis**
  The system can detect which products or categories have higher return rates and identify the reasons behind them. This helps improve product quality and reduce return losses.

REFERENCES

Journal

[1] Kaggle – Brazilian E-commerce Public Dataset by Olist (used for real-world sales, customers, and delivery data).

[2] Aurélien Géron – *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow* (for ML concepts and implementation)

[3] Choudhary, S., & Shukla, S. (2020). "Sales Forecasting for E-commerce Using Machine Learning Techniques" — *International Journal of Scientific Research in Computer Science and Engineering (IJSRCSE)*.

[4] Tsai, C. F., & Hung, C. S. (2019). "Customer Purchase Behavior Prediction in E-commerce using Data Mining Techniques"

GitHub ID:

https://github.com/hariiom08/Major-1.git

Guided By:

Prof. Mahendra Verma

Group Members:

Shubham Siloriya (0827CD221068)

Deepak Patel (0827CD221023)

Hariom Mandloi (0827CD221033)

Deepak Patel (0827CD221024)