MACHINE LEARNING

Project Report

FALLSEM 2024-25

Title: E-COMMERCE SHIPMENT PREDICTION USING MACHINE LEARNING

Submitted by

K.PRADEEP (22BCE9408)

Guide:
Prof.Beebi Naseeba

VIT-APUNIVERSITY

AMARAVATI

ANDHRAPRADESH, INDIA

2024

INTRODUCTION :

In the fast-paced world of e-commerce, where customer expectations for speedy delivery are higher than ever, accurate shipment prediction has become a critical challenge. This project aims to develop a robust machine learning model that can accurately forecast delivery times, enabling e-commerce companies to optimize logistics, enhance customer satisfaction, and improve operational efficiency E-commerce has seen exponential growth, necessitating efficient shipment prediction systems to enhance customer satisfaction and optimize logistics. This project aims to develop a machine learning model that predicts shipment delivery times based on historical data. By leveraging various predictive algorithms, the model aims to improve accuracy in delivery estimates, allowing businesses to manage logistics more effectively. The integration of external factors such as weather, traffic, and regional holidays will also be explored. This project contributes to the body of knowledge in e-commerce logistics by providing a framework that can be adapted and scaled for various operational needs.

PROJECT SCOPE :

The "E-commerce Shipment Prediction" project aims to develop a system that predicts delivery times, shipping costs, and potential delays for e-commerce shipments. Key features include forecasting delivery times based on location, carrier, and external factors like weather, estimating shipping costs, and identifying potential delays or inventory needs. Data sources will include internal e-commerce order and shipping data, as well as external data from carriers, weather services, and public holidays. The deliverables include a data pipeline, predictive models, integration with e-commerce platforms, and a dashboard for shipment tracking and predictions. The project timeline is estimated at 2-6 months, with success metrics focused on achieving 85% prediction accuracy, improving delivery reliability, and reducing shipping costs. The technology stack will involve Python, machine learning frameworks like Scikit-learn, and integration with APIs for real-time data. Risks include data quality and scalability, mitigated by thorough testing, validation, and cloud-based infrastructure..

SOFTWARE COMPONENTS :

- SPYDER

- HTML(Hyper Text Mark-up Language)
- CSS(Cascadian Style Sheets)
- FLASKING
- PYTHON

PROPOSED APPROACHES :

The proposed approach utilizes advanced machine learning algorithms to enhance shipment time predictions. The process will include the following steps:

- **Data Collection** : Aggregate historical shipment data, including delivery times, locations, product types, and external factors.
- **Feature Engineering** : Identify and create features that may influence delivery times, such as weather conditions, traffic patterns, and customer location.
- **Model Selection** : Experiment with various machine learning models, including Decision Trees, Random Forests, Gradient Boosting, and Neural Networks.
- **Model Evaluation** : Use metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to evaluate model performance. Crossvalidation techniques will be employed to ensure robustness.
- **Integration of External Factors** : Incorporate real-time data such as traffic and weather forecasts to adjust predictions dynamically.
- **User Interface** : Develop a dashboard for businesses to visualize predicted delivery times and factors influencing them.

EXISTED APPROACHES :

- **Linear Regression**: Used when there's a linear relationship between variables like shipment size, distance, and delivery time.

- **Random Forest / Gradient Boosting Machines (GBM)**: More effective for capturing non-linear relationships, particularly when there are many features.

•**Neural Networks**: Can capture complex patterns, especially when there are multiple influencing factors (e.g., weather, delivery route).

- **Time Series Forecasting (ARIMA, Prophet)**: Used for forecasting delivery times based on historical data and trends, especially for long-term predictions.

- **Key Features :**

•**Order Size**: Large orders may take longer.

•**Shipping Method**:
Different shipping options (e.g., expedited shipping, standard shipping) will affect delivery time.

•**Customer Location**: Distance from the warehouse and infrastructure in the destination area (urban vs. rural) can be important.

•**Warehouse Location**: Proximity to the destination.

•**Historical Shipping Data**: Previous delivery times for similar orders or customers.

How the proposed model will overcome the real-world problems?

By this model we can improve the statistics and also the efficiency of the device to overcome the problem in a very efficient way. So, we can identify the shipment prediction of each product by entering the details and get respective output.

Objectives:

The **objectives of E-commerce Shipment Prediction** are aimed at improving the efficiency and reliability of the shipping process, ultimately enhancing both operational performance and customer satisfaction. Key objectives include:

1. **Accurate Delivery Time Predictions**: Forecast delivery times based on factors like order processing time, shipping method, geographic location, and potential disruptions (e.g., weather or traffic), providing customers with reliable delivery windows.

2. **Optimizing Shipping Costs**: Estimate shipping costs for different delivery methods and package weights, allowing businesses to offer cost-effective shipping options while maintaining profitability.

3. **Predicting Shipment Delays**: Identify potential delays in advance based on historical patterns and external factors, enabling proactive communication with customers and improving customer service.

4. **Demand Forecasting for Inventory Management**: Predict shipping demand based on historical sales data, allowing businesses to better manage inventory levels and reduce stockouts or overstocking.

5. **Enhancing Customer Experience**: Provide customers with real-time, accurate delivery estimates, reducing uncertainty and increasing satisfaction through transparency.

6. **Operational Efficiency**: Streamline logistics and fulfillment processes by providing businesses with better insights into shipment timelines, reducing operational costs, and improving resource allocation.

7. **Data-Driven Decision Making**: Use predictive analytics to make informed decisions about shipping policies, carrier selection, and inventory management, optimizing overall business operations.

PROCESSING THROUGH SO MANY MODELS LIKE :

- GRADIENT BOOSTING
- RANDOM FOREST
- SUPPORT VECTOR MACHINE

- KNN-ALGORITHM

MODEL OBJECTIVE :

1. Gradient Boosting :

Use Case: Can be used for both classification (e.g., will the shipment arrive on time) and regression (e.g., how long will the shipment take).
•**Typical Accuracy:** Depending on the data and fine-tuning. These models are generally more accurate than decision trees and random forests in many cases.
•**Pros:** Excellent performance, handles missing data well, robust to overfitting when tuned.
•**Cons:** More complex, requires careful hyperparameter tuning.

2. Random Forests :

•**Use Case:** Similar to decision trees, but an ensemble method that can predict more accurately by reducing overfitting.
•**Typical Accuracy:** Random Forests tend to perform well on shipment prediction tasks, especially when there is enough data and features.
•**Pros:** Handles large datasets well, reduces overfitting, robust.
•**Cons:** Less interpretable than decision trees.

3. Support Vector Machines (SVM) :

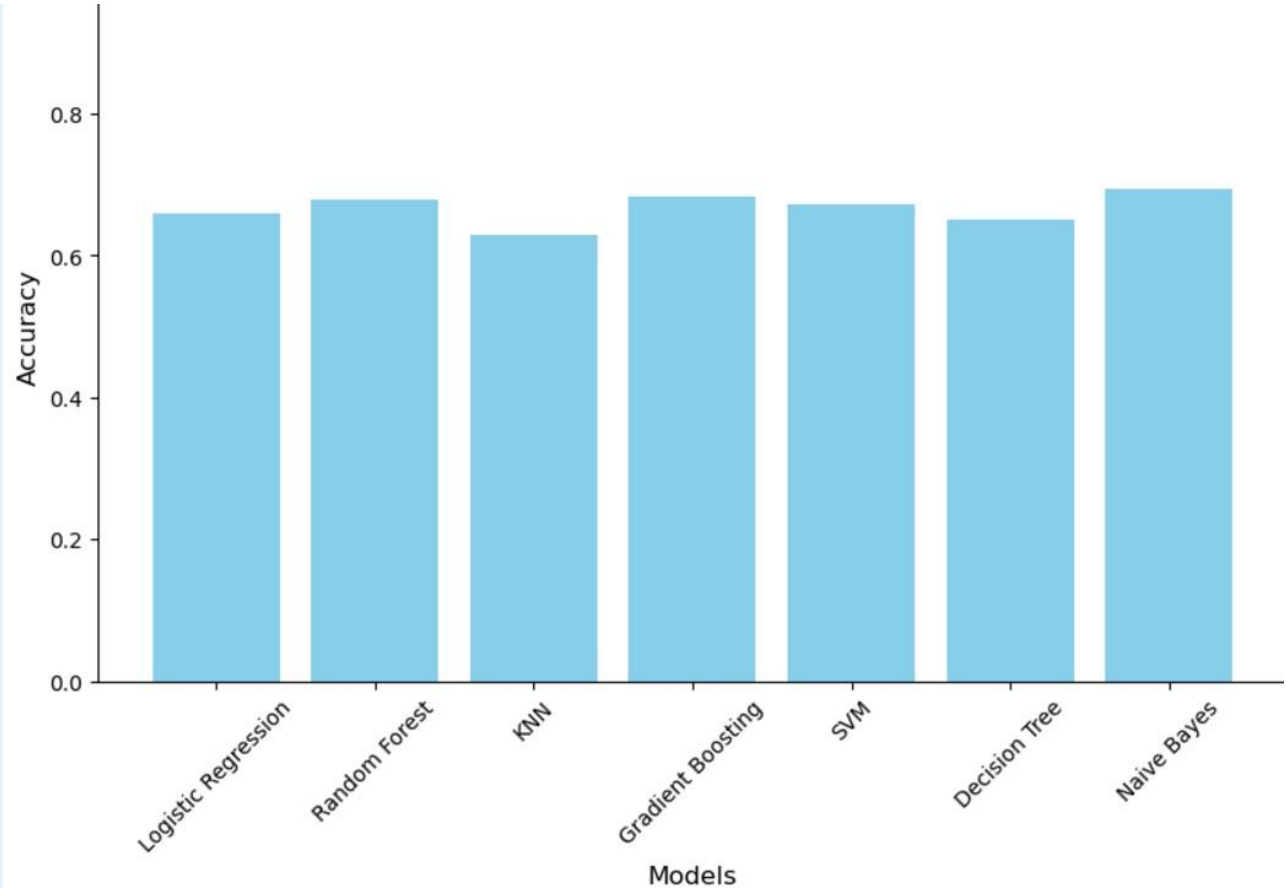
•**Use Case:** Binary classification tasks like predicting whether a shipment will arrive on time.
•**Typical Accuracy:** SVM can work well for classification tasks with clear margins between classes, but may struggle with large datasets.
•**Pros:** Effective in high-dimensional spaces.
•**Cons:** Requires careful tuning, can be computationally expensive.

4. K-Nearest Neighbors (KNN) :

•**Use Case:** Classification tasks, such as predicting whether a shipment will be delayed based on features like distance, time of year, etc.
•**Typical Accuracy:** Accuracy depends heavily on the choice of "k" and feature scaling.

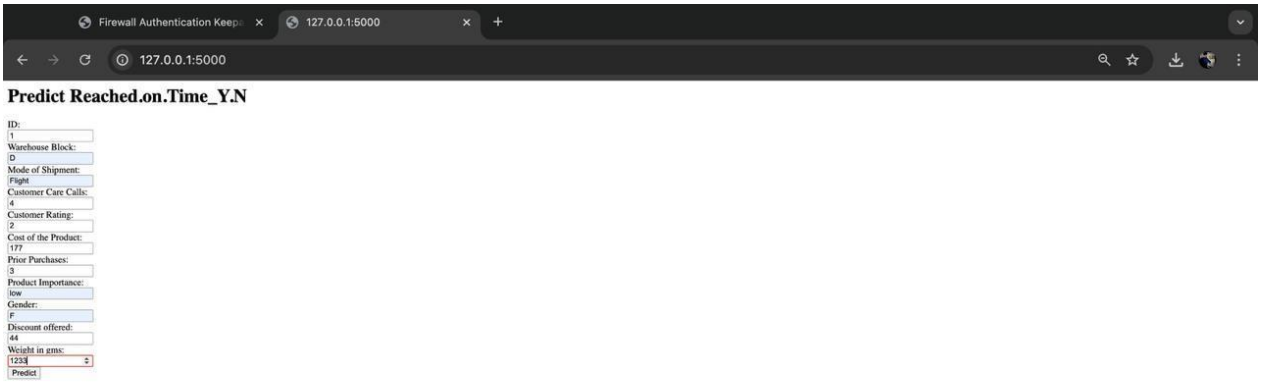
•**Pros:** Simple, works well with smaller datasets.
•**Cons:** Computationally expensive for large datasets, sensitive to the choice of features.

ACCURACY OF THE ABOVE MODELS :



STEPS : INPUT

:



OUTPUT :

CODE : Execution in SPYDER from ANACONDA

- app.py
- app.py.html
- index.html

DATASET :

- https://github.com/pradeep9408/E-COMMERCE-SHIPMENT-PREDICTION-USING-MACHINE-LEARNING/blob/main/Train.csv

FUNCTIONS :

- The code is executed in python programming language.
- The functionality is done through the SPYDER.
- Therefore by executing it provides the IP address and paste the address in the browser to enter the web page and proceed further.

CONCLUSION :

- After looking into many machine learning models, we found that Gradient boosting is well suitable for the E-commerce shipment prediction.
- Apart from Gradient boosting, analyzing the accuracy and prediction rate, RANDOM FOREST algorithm is also best suitable for the project.
- Therefore, as it well known for handling large datasets and have faster rate of accuracy and prediction rate, we prefer the RANDOM FOREST algorithm.

THANK YOU