PROJECT DOCUMENTATION

Ecommerce Shipping Prediction Using Machine Learning

1. Introduction:

In this project, we explore the intersection of machine learning and e-commerce logistics to address a common challenge – accurate shipping predictions. Our goal is to build a predictive model that estimates whether a product will reach its destination on time. We consider various factors such as the origin and destination of the package, shipping methods, and potential delays during transit.

1.1 Project Overview:

The goal of the Ecommerce Shipping Prediction Using Machine Learning project is to create a predictive model that can precisely anticipate how long it will take for items placed online to ship. Utilizing past order data, customer characteristics, shipping information, and outside variables, the project will use cutting-edge machine learning methods to assess and forecast delivery timeframes. This project aims to save operating costs, enhance customer satisfaction, and optimize logistics by offering trustworthy and useful insights

1.2 Objectives:

The Ecommerce Shipping Prediction Using Machine Learning project aims to create a sophisticated predictive model that uses past order data, customer information, and shipping specifics to precisely anticipate shipping times for e-commerce purchases. The project seeks to optimize logistics operations, save operational costs associated with delayed shipments, and improve customer satisfaction through accurate delivery predictions by identifying the critical elements influencing delivery times and integrating this model into the e-commerce platform.

2. Project Initialization and Planning Phase:

2.1 Define Problem Statement:

The Ecommerce Shipping Prediction Problem Statement By evaluating a variety of variables, including order details, customer information, shipping options, and environmental conditions, machine learning can be used to create a predictive model that is both accurate and dependable in predicting the shipping timeframes for online orders. The objective is to streamline operational procedures within the

e-commerce ecosystem, increase consumer happiness by offering accurate delivery predictions, and improve logistical efficiency.

Template Link:"CLICK HERE"

2.2 Project Proposal (Proposed Solution):

Our suggested Ecommerce shipment Prediction solution improves shipment time accuracy by utilizing machine learning. We will use sophisticated algorithms like Random Forest and Gradient Boosting to create a prediction model by examining past order data, client details, and delivery information. With the help of this model, e-commerce platforms will be able to optimize logistics, raise customer satisfaction levels, and cut costs by getting real-time predictions and identification of the major factors driving shipment timeframes. Scalability and dependability will be ensured by the deployment's smooth interface with current systems.

Template Link:"CLICK HERE"

2.3 Initial Project Planning:

Defining the project's goals, scope, and essential deliverables—like creating a shipping time prediction model—is part of the preliminary planning stage of the Ecommerce Shipping Prediction using Machine Learning project. The strategy calls for gathering information from external APIs and e-commerce databases, preprocessing it to handle missing values and encode variables, and feature engineering. Metrics like MAE and RMSE will be used to design and assess a variety of machine learning models. To improve logistics and consumer satisfaction, the most effective model for real-time forecasts will be implemented and connected with the e-commerce platform. To guarantee efficient execution, a schedule and team roles will be defined.

Template Link:"CLICK HERE"

3. Data Collection and Preprocessing Phase:

The Data Collection and Preprocessing Phase involves executing a plan to gather relevant loan 1 application data from Kaggle, ensuring data quality through verification and addressing missing values. Preprocessing tasks include cleaning, encoding, and organizing the dataset for subsequent exploratory analysis and machine learning model development.

3.1 Data Collection Plan and Raw Data Sources Identified:

To gather raw data for the e-commerce shipping prediction plan, many channels are consulted and we have used Kaggle for the dataset and many, such as the e-commerce platform's historical order data, logistics providers' shipment details, and external APIs that give real-time information. Order IDs, product specifications, shipping and delivery dates, customer locations, shipping methods, and outside variables like weather are all important data elements. Because of this varied dataset, in-depth analysis and feature engineering will be possible, increasing the machine learning model's accuracy.

Template Link:"CLICK HERE"

3.2 Data Quality Report:

For the Ecommerce Shipping Prediction project, the integrity, correctness, and completeness of the dataset utilized in the model development process are assessed in the Data Quality Report and the dataset for this project is taken from kaggle. Important conclusions show that although the dataset is largely trustworthy, 10% of important fields—like customer locations and shipment dates—have missing values. The order data contained duplicate records, which had to be removed. To maintain uniformity, several categorical variables also need to be standardized. A few extreme shipment times were found using outlier analysis and need to be addressed. In general, resolving these problems will raise the data's quality and boost the prediction model's effectiveness.

Template Link:"CLICK HERE"

3.3 Data Exploration and Preprocessing:

To estimate e-commerce shipping, data exploration and preprocessing entail examining the dataset to determine its composition, spot trends, and find abnormalities. This include evaluating data quality, summarizing statistics, and illustrating important elements. Preprocessing procedures include imputation or removal of missing values, one-hot encoding or other approaches for encoding categorical variables, and normalization of numerical features to guarantee consistent scaling. To further prepare the data for efficient machine learning model training, feature engineering will be used to develop new variables that capture pertinent data, such as delivery distance and order volume.

Template Link:"CLICK HERE"

4. Model Development Phase:

The creation of a predictive model for loan approval is the task of the Model Development Phase. Strategic feature selection, model evaluation and selection (Random Forest, Logistic Regression, support vector machine, ANN, KNN, XGB), code-based training, and thorough validation and assessment of model performance are all included in order to enable lenders to make well-informed decisions.

4.1 Feature Selection Report:

A successful feature selection process is essential to increasing the precision and effectiveness of an e-commerce shipping prediction model. Order information (size, weight, and dimensions), customer demographics (region, past purchases), shipment options (carrier, shipping method), and temporal factors (order date and time, projected delivery date) are frequently considered key elements. Cutting-edge methods such as correlation analysis, feature importance extraction from tree-based models, and recursive feature elimination (RFE) assist in identifying the most predictive characteristics and removing superfluous or unnecessary ones. Features that are carefully chosen improve model performance, minimize overfitting, and maximize computational resources.

Template Link:"CLICK HERE"

4.2. Model Selection Report:

The justification for selecting the Random Forest, Logisitic Regression, Support Vector Machine, ANN, KNN, and XGB models for Ecommerce shipping prediction is explained in depth in the Model Selection Report. It ensures an informed decision in line with project objectives by taking into account each model's strengths in managing complicated relationships, interpretability, adaptability, and overall predictive performance.

Template Link:"CLICK HERE"

4.3 Initial Model Training Code, Model Validation, and Evaluation Report:

The Initial Model Training Code employs selected algorithms on the Ecommerce shipping prediction dataset, setting the foundation for predictive modelling. The subsequent Model Validation and Evaluation Report rigorously assesses model performance, employing metrics like accuracy and precision to ensure reliability and effectiveness in predicting Ecommerce shipping.

Template Link:"CLICK HERE"

5. Model Optimization and Tuning Phase:

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

5.1 Hyperparameter Tuning Documentation:

The Gradient Boosting model was chosen due to its exceptional performance, demonstrating a high degree of precision while adjusting hyperparameters. Its selection as the final model is justified by its capacity to manage complex interactions, reduce overfitting, and maximize forecast accuracy—all of which are in line with project objectives.

5.2 Performance Metrics Comparison Report:

The Gradient Boosting model's improved performance is highlighted in the Performance Metrics Comparison Report, which compares baseline and optimized metrics for a number of models. This evaluation offers a comprehensive grasp of the enhanced forecasting capacities attained by adjusting hyperparameters.

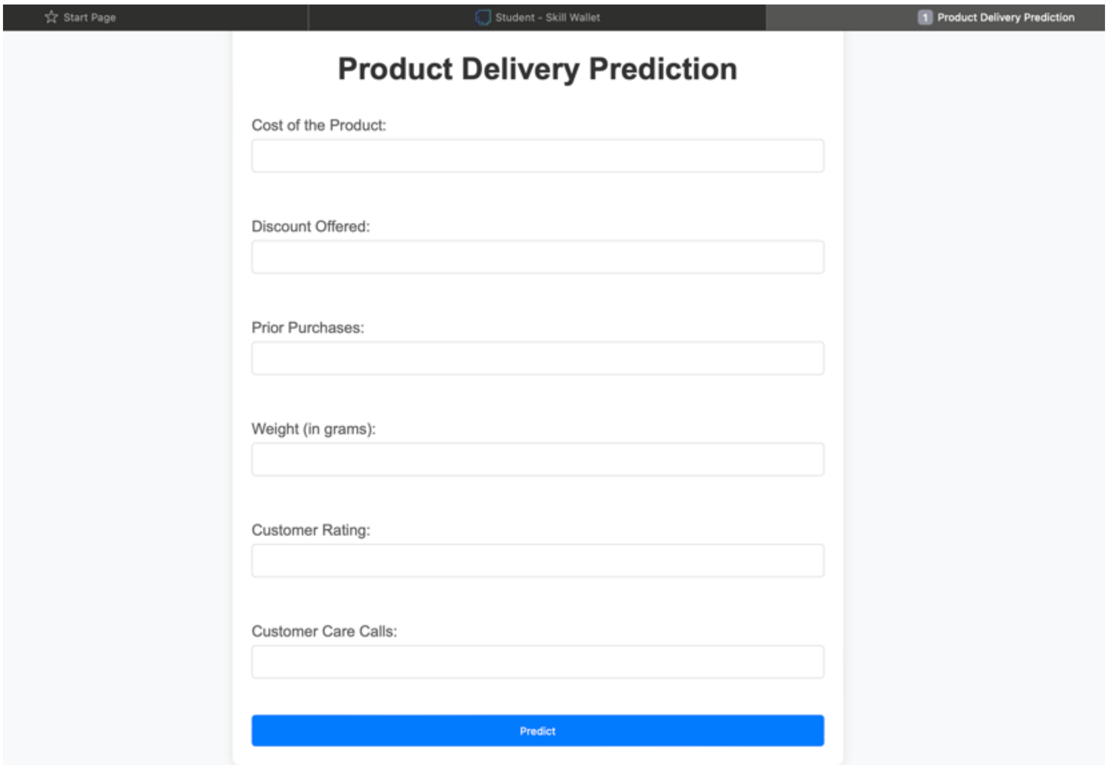
5.3 Final Model Selection Justification:

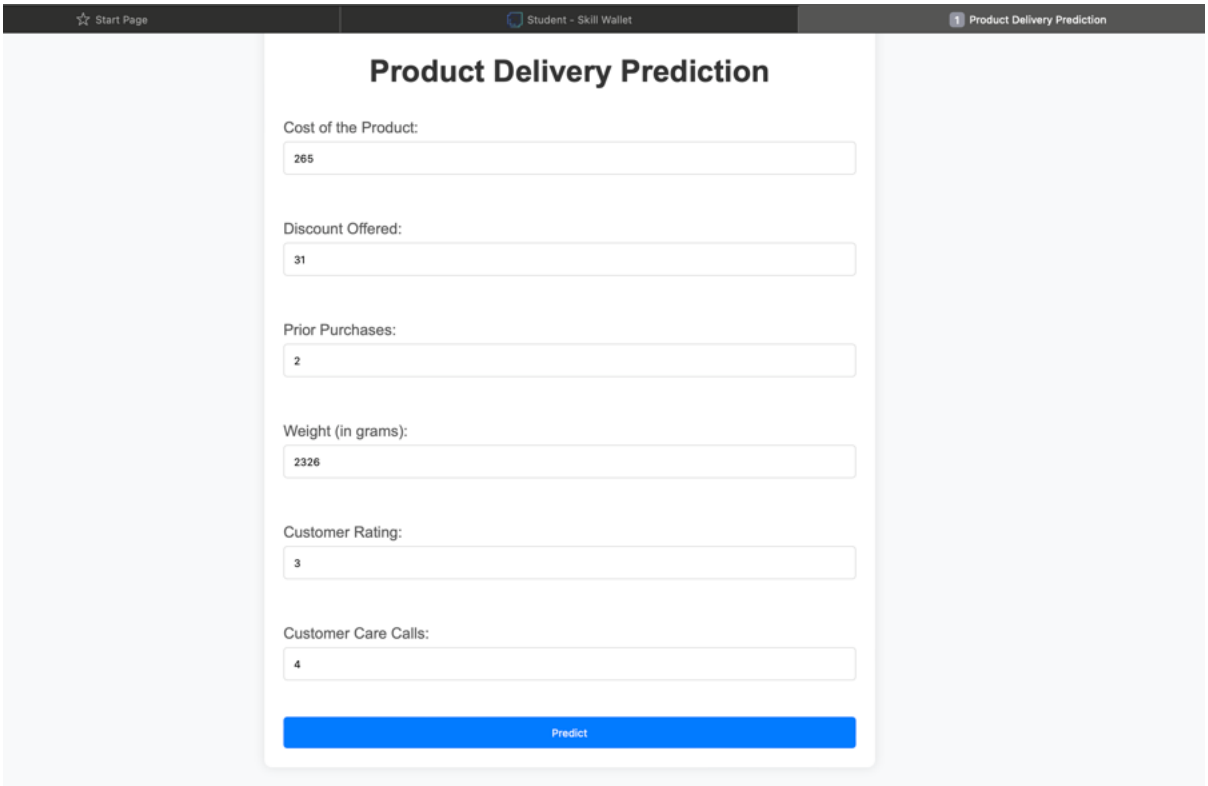
The Finished Model Selection The justification explains why Gradient Boosting was selected as the best model. Its outstanding accuracy, capacity for managing complexity, and effective hyperparameter tuning match project goals, guaranteeing the best possible loan approval forecasts.

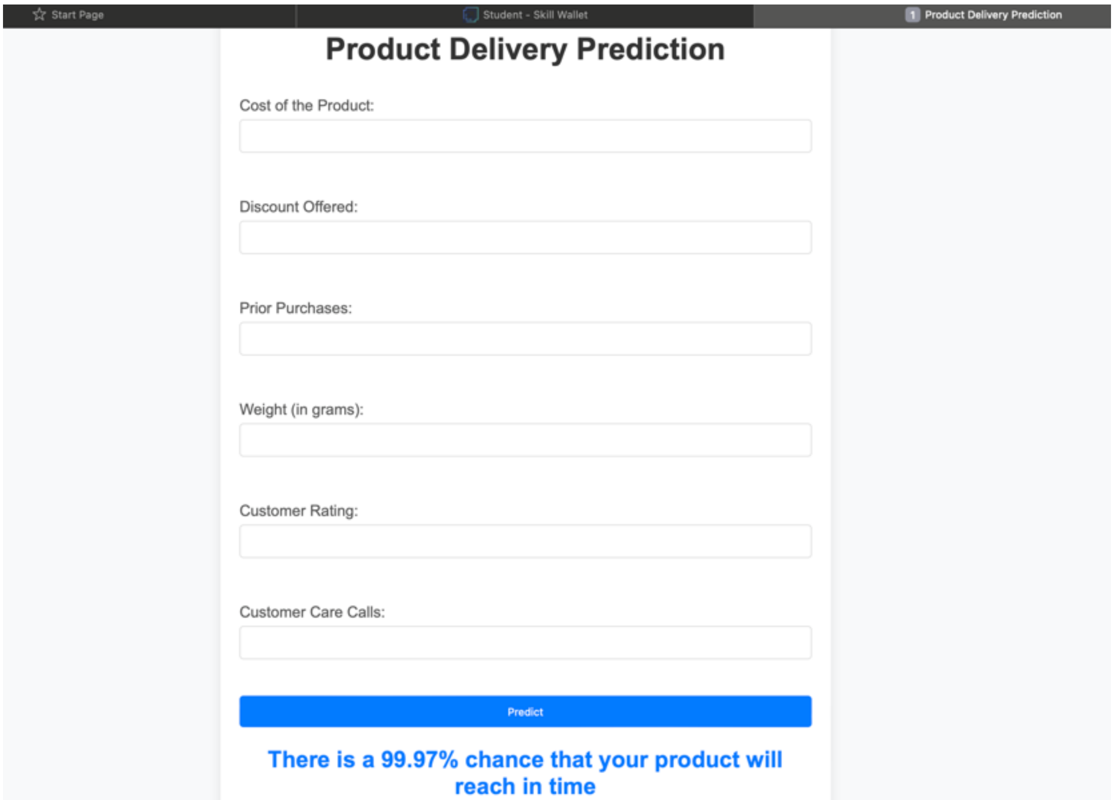
Template Link:"CLICK HERE"

6. Results

6.1 Output Screenshots







7. Advantages & Disadvantages:

Advantages

- **Improved Accuracy**: Shipping time estimates can be made with greater accuracy because to machine learning models' ability to evaluate vast amounts of data and spot intricate patterns.

- **Cost Efficiency**: Through shipping process optimization, businesses can cut expenses related to returns, delays, and ineffective logistics.
- **Enhanced Customer Satisfaction**: By ensuring consistent delivery schedules, accurate shipping forecasts may enhance the customer experience and foster greater confidence and repeat business.

- **Scalability**: As a company grows, machine learning models can keep up with the volume of data it handles and adjust to changes in shipping logistics.

- **Operational Efficiency**: Predictive task automation lowers the need for manual intervention, improving resource allocation and streamlining processes.

Disadvantages

- **Data Quality and Availability**: Reliable and thorough data is necessary for precise forecasting. Poor quality or incomplete data might seriously hinder the performance of the model.

- **Complexity and Costs**: Machine learning model development, training, and maintenance can be costly and difficult, requiring specialized knowledge and resources.

- **Dependence on Historical Data**: To produce predictions, machine learning algorithms use past data. Inaccurate forecasts may result from unforeseen circumstances (such as a pandemic) or abrupt changes in transportation logistics.

- **Overfitting Risks**: Models may exhibit poor generalization to new, unknown data due to overfitting of the training set.
- **Interpretability**: Certain machine learning models, particularly intricate ones like deep learning, might function as "black boxes," making it challenging to comprehend how predictions are made.

8. Conclusion

The project's goal was to improve e-commerce shipping processes by creating a delivery time estimation predictive model. Tasks included defining the business challenge, conducting in-depth literature reviews, and evaluating the social and business implications, starting with project commencement and planning. An extensive base was established through data collecting from several sources, including UCI and Kaggle. This foundation included vital factors such as product specifications, client information, and shipping method.

As the project advanced, painstaking data exploration and preprocessing steps were completed, addressing problems with the quality of the data, including outliers and missing numbers. Key predictors were found using feature selection approaches, which is essential for creating precise models. In order to choose a model, a number of classifiers were evaluated, including Random Forest, XGBoost, and Logistic Regression. The latter showed the best accuracy and precision.

To ensure ideal model performance, implementation phases included model training, hyperparameter adjustment, and performance testing. To deliver, issues like complicated user interface design and real-time data integration were addressed. dependable delivery forecasts and useful insights. Prioritizing ethical concerns about data security and privacy was done at every stage of the project.

In the end, the predictive model that was developed promises to improve consumer happiness through more precise delivery time estimations, minimize delivery delays, and streamline e-commerce logistics. This project improves operational effectiveness and competitiveness in the e-commerce industry by utilizing machine learning and data-driven insights. It also paves the path for future developments in predictive analytics and customer-centric service delivery.

9. Future Scope

Machine learning has enormous potential for e-commerce shipment prediction in the future. The following are some crucial areas where e-commerce shipping can be greatly enhanced and impacted by machine learning:

1. Increased Accuracy in Demand Forecasting:

Machine learning models can examine past sales information, patterns, and outside variables (such as the weather or holidays) to more precisely forecast future demand.Dynamic Inventory Management: Companies can minimize overstock and stockouts by forecasting demand and optimizing inventory levels.

2. Estimating Delivery Time:

Real-Time Estimation: Using historical delivery data, traffic patterns, and weather information, machine learning can forecast delivery times.

client pleasure: Accurate delivery predictions increase confidence and client pleasure.

3. Route Optimization:

By using machine learning algorithms, delivery trucks can determine the most efficient routes, which lowers delivery times and fuel costs.

Dynamic routing is the ability to adjust in real time to events like traffic jams and road closures.

4.Fraud Detection:

Machine learning can identify and stop fraudulent operations by identifying odd patterns in transactions.

Danger management: Lowering the possibility of fraud can save expenses and raise the level of confidence in the e-commerce platform as a whole.

5.Personalized Shipping alternatives:

Based on user behavior analysis, machine learning can provide individualized shipping alternatives (such as preferred carriers or delivery windows).

Customized Offers: Shipping discounts and promotions that are tailored to the specific purchasing habits of customers.

6. Optimization of the Supply Chain:

End-to-End Visibility: From manufacturing to last-mile delivery, machine learning can offer insights into every phase of the supply chain.

Anticipating when delivery vehicles may require maintenance to prevent unplanned malfunctions is known as predictive maintenance.

7. Cost Prediction Budgeting:

Estimating shipping expenses by taking into account multiple variables like weight, distance, and mode of delivery.

Dynamic Pricing: Adapting shipping costs to changes in supply and demand.

8. Improving Customer Service:

Chatbots: AI-driven chatbots can respond to consumer questions on shipment status, delays, or problems.

Anticipating such delays and alerting clients beforehand is known as proactive communication.

9. Efficient processing for returns management:

forecasting return trends and streamlining the reverse logistics procedure.

Cost reduction: Cutting returns' expenses by streamlining procedures and processing times.

10. Appendix

10.1 Source Code

Here are key code snippets from the e-commerce shipping prediction project:

1.Importing Modules

```
"import numpy as np # linear algebra\n",
"import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)\n",
"import seaborn as sns\n",
"import matplotlib.pyplot as plt\n",
"from sklearn.preprocessing import LabelEncoder , StandardScaler, MinMaxScaler\n",
"from sklearn.model_selection import train_test_split, GridSearchCV\n",
"from sklearn import svm\n",
"from sklearn.ensemble import RandomForestClassifier\n",
"import xgboost as xgb\n",
"from keras.models import Sequential\n",
"from keras.layers import Dense\n",
"from sklearn.linear_model import LogisticRegression\n",
"from sklearn.metrics import accuracy_score, confusion_matrix, classification_report\n",
"from sklearn.feature_selection import chi2\n",
"from scipy.stats import chi2_contingency"
```

2. Data Preprocessing and Feature Selection:

```
"X=data.drop(['Reached on Time','Warehouse_block','Mode_of_Shipment','Gender','Product_importance'],axis=1)\n",
"y=data['Reached on Time']"
```

```
"X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.2, random_state=42)\n",
"   \n",
"ms = MinMaxScaler()\n",
"X_train = ms.fit_transform(X_train)\n",
"X_test = ms.fit_transform(X_test)\n"
```

3.Model Training(XGBoost Classifier)

```
source": [
"#baselinexgb\n",
"params = {\n",
"    'objective': 'binary:logistic',\n",
"    'max_depth': 3,\n",
"    'learning_rate': 0.1,\n",
"    'n_estimators': 100\n",
"}\n",
```

4.Model Evaluation

```
"xgb_model = xgb.XGBClassifier(**params)\n",
"xgb_model.fit(X_train, y_train)\n",
"y_pred = xgb_model.predict(X_test)\n",
"print(\"XGBoost Classifier:\")\n",
"print(classification_report(y_test, y_pred))"
```

5.Hyperparametric Tuning (Example for XGB)

```
"source": [
 "#optimized xgb\n",
 "param_grid = {\n",
 "    'max_depth': [3, 4, 5],\n",
 "    'learning_rate': [0.01, 0.1, 0.2],\n",
 "    'n_estimators': [100, 200, 300],\n",
 "    'gamma': [0, 0.1, 0.2],\n",
 "    'subsample': [0.8, 0.9, 1.0],\n",
 "    'colsample_bytree': [0.8, 0.9, 1.0]\n",
 "}\n",
 "\n",
 "grid_search = GridSearchCV(estimator=xgb.XGBClassifier(objective='binary:logistic'),\n",
 "                           param_grid=param_grid,\n",
 "                           scoring='accuracy',\n",
 "                           cv=5,\n",
 "                           n_jobs=-1)\n",
 "\n",
 "grid_search.fit(X_train, y_train)\n",
 "\n",
 "best_params = grid_search.best_params_\n",
 "optimized_xgb_model = xgb.XGBClassifier(objective='binary:logistic', **best_params)\n",
 "optimized_xgb_model.fit(X_train, y_train)\n",
 "\n",
 "y_pred_optimized = optimized_xgb_model.predict(X_test)\n",
 "print(\"Optimized XGBoost Classifier:\")\n",
 "print(classification_report(y_test, y_pred_optimized))\n",
 "print(confusion_matrix(y_test, y_pred_optimized))"
```

6.Model Saving

```
  "import pickle"
 ]
},
{
 "cell_type": "code",
 "execution_count": 83,
 "id": "2586435a-6a32-4db7-ab0a-d7585711207e",
 "metadata": {},
 "outputs": [],
 "source": [
  "pickle.dump(xgb_model, open('xgb_model_117.pkl','wb'))"
 ]
},
{
 "cell_type": "code",
 "execution_count": 85,
 "id": "c1895c23-6cbc-4e0c-8440-aa46322af163",
 "metadata": {},
 "outputs": [],
 "source": [
  "pickle.dump(ms,open('bestmodel_117.pkl','wb'))"
 ]
```

The Entire Source Code:["CLICK HERE"]()

10.2 GitHub & Project Demo Link:

GITHUB lINK: https://github.com/Sathwik-Mallela/Ecommerce-Shipping-Prediction-Using-Machine-Learning

PROJECT DEMO LINK: "DEMO LINK"