Ecommerce Shipping Prediction using Machine Learning

Introduction:

1.1. Project Background

The rise of ecommerce has significantly increased the volume of shipped goods worldwide. Accurate prediction of shipping times is crucial for enhancing customer satisfaction and optimizing logistics operations.

1.2. Project Overview

This project aims to develop a machine learning model to predict the shipping duration for ecommerce orders based on various factors such as order details, shipping methods, and historical data.

Project Overviews

Objectives

1. **Primary Objective:**

   Develop a machine learning model to accurately predict shipping times.

2. **Secondary Objectives:**

   Analyse the factors affecting shipping times.

   Improve logistics planning and customer satisfaction through better shipping time estimates.

Project Initialization and Planning Phase

2.1. Define Problem Statement

Online shoppers frequently encounter uncertainty and delays in receiving their ordered products due to unreliable shipping estimates provided by ecommerce platforms. This inconsistency often leads to frustration and dissatisfaction among customers, impacting their overall shopping experience negatively. Customers desire accurate and reliable predictions of shipping times at the point of purchase to make informed decisions and manage expectations effectively. Therefore, there is a critical need to develop a machine learning solution that can predict shipping times accurately based on various factors such as product availability, shipping location, and historical shipping data. By addressing this challenge, ecommerce platforms can enhance customer satisfaction, improve trust in their services, and ultimately increase repeat purchases.

2.2. Project Proposal (Proposed Solution)

The proposed project, “Ecommerce Shipping Prediction suing Machine Learning” aims to leverage machine learning for more accurate Ecommerce Shipping predictions. Using a comprehensive dataset including (12) id, warehouse block, mode of shipment, customer care calls, customer ratings, cost of products, prior purchases, product importance., gender, discount offered, weight, time of arrival. The project seeks to develop a predictive model predicting the Shipment of all the Ecommerce products.

2.3. Initial Project Planning

Timeline:

This section contains all the milestones and the deadlines.
Milestones contain the following stages:

Milestone-1: Project Initialisation and Planning Status
Milestone-2: Data Collection and Preprocessing Phase
Milestone-3: Model Development Phase
Milestone-4: Model Optimisation and Tuning Phase
Milestone-5: Project Executable Files

Resources:

The resources for this project and dataset were provided and saved in the Kaggle platform where we got a templates and references for making this Ecommerce Shipment Prediction using Machine Learning .

Risk Assessment:

Developing an E-Commerce Shipping prediction model using machine learning involves various risks. Here are some of the key risks:

1. **Data Quality and Availability:**
    - **Incomplete or Inaccurate Data:** Incomplete, inaccurate, or outdated data can lead to unreliable predictions.
    - **Data Bias:** Biased data can result in biased models, affecting fairness and accuracy.
2. **Model Performance:**
    - **Overfitting/Underfitting:** Overfitting to training data or underfitting can result in poor generalization to new data.
    - **Model Drift:** Changes in shipping patterns over time may cause the model to degrade, necessitating frequent retraining.
3. **Implementation Risks:**
    - **Integration Issues:** Integrating the model into existing systems can be complex and may cause disruptions if not managed properly.
    - **Scalability:** Ensuring the model can handle large volumes of data and transactions efficiently is crucial.
4. **Security and Privacy:**
    - **Data Breaches:** Protecting sensitive customer and shipping data from unauthorized access is critical.
    - **Compliance:** Adhering to data protection regulations (e.g., GDPR) is necessary to avoid legal issues.
5. **Ethical Concerns:**
    - **Bias and Fairness:** Ensuring the model does not unfairly disadvantage certain groups or individuals.
    - **Transparency:** Providing clear explanations for predictions to maintain trust and accountability.
6. **Operational Risks:**
    - **Reliability:** Ensuring the model operates reliably in real-time environments.
    - **Maintenance:** Ongoing maintenance and monitoring are required to ensure the model continues to perform well.
7. **Economic Risks:**
    - **Cost:** The cost of developing, deploying, and maintaining the model can be significant.
    - **ROI:** There is a risk that the investment in the model may not yield the expected return on investment.
8. **Stakeholder Alignment:**
    - **Expectations Management:** Aligning the expectations of various stakeholders (management, customers, etc.) with the model's capabilities and limitations.

Data Collection and Preprocessing Phase

3.1. Data Collection Plan and Raw Data Sources Identified

Data Sources:

The dataset for "Ecommerce Shipping Prediction using Machine Learning" is sourced from Kaggle. It includes many details such as id, warehouse block, mode of shipment, customer care calls, customer ratings, cost of products, prior purchases, product importance., gender, discount offered, weight, time of arrival. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

3.2. Data Collection

The proposed project, "Ecommerce Shipping Prediction suing Machine Learning" aims to leverage machine learning for more accurate Ecommerce Shipping predictions. Using a comprehensive dataset including id, warehouse block, mode of shipment, customer care calls, customer ratings, cost of products, prior purchases, product importance, gender, discount offered, weight, time of arrival. The project seeks to develop a predictive model predicting the Shipment of all the Ecommerce products.

Ecommerce Data Collection Report: Click Here

3.3. Data Quality Report

The dataset for " Ecommerce Shipping Prediction suing Machine Learning " is sourced from Kaggle. It includes comprehensive dataset including id, warehouse block, mode of shipment, customer care calls, customer ratings, cost of products, prior purchases, product importance, gender, discount offered, weight, time of arrival. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

Ecommerce Data Quality Report: Click Here

3.4. Data Exploration and Preprocessing

Data Exploration involves analyzing given dataset to understand patterns, distributions, and outliers. Preprocessing includes handling missing values, scaling, and encoding categorical variables. These crucial steps enhance data quality, ensuring the reliability and effectiveness of subsequent analyses in the loan approval project.

Data Exploration and Preprocessing Report: Click Here

Model Development Phase

4.1. Feature Selection Report

The Feature Selection Report outlines the rationale behind choosing specific features for the prediction of the Shipment arrival .It evaluates relevance, importance, and impact on predictive accuracy, ensuring the inclusion of key factors influencing the model's ability to discern credible loan applicants.

Feature Selection Report: Click Here

4.2. Model Selection Report

The Model Selection Report details the rationale behind choosing Random Forest, Decision Tree, KNN, and XGB models for loan approval prediction. It considers each model's strengths in handling complex relationships, interpretability, adaptability, and overall predictive performance, ensuring an informed choice aligned with project objectives.

Model Selection Report: Click Here

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

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

Model Development Phase Template: Click Here

Model Optimization and Tuning Phase

5.1. Hyperparameter Tuning Documentation

The Gradient Boosting model was selected for its superior performance, exhibiting high accuracy during hyperparameter tuning. Its ability to handle complex relationships, minimize overfitting, and optimize predictive accuracy aligns with project objectives, justifying its selection as the final model

5.2. Performance Metrics Comparison Report

The Performance Metrics Comparison Report contrasts the baseline and optimized metrics for various models, specifically highlighting the enhanced performance of the Gradient Boosting model. This assessment provides a clear understanding of the refined predictive capabilities achieved through hyperparameter tuning.

5.3. Final Model Selection Justification

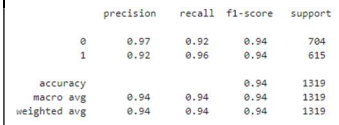
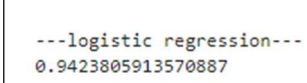
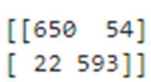
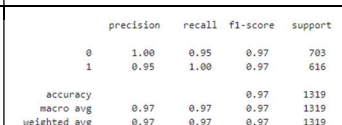
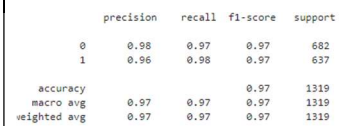
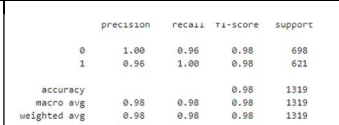
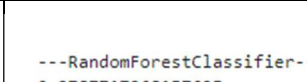
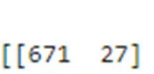
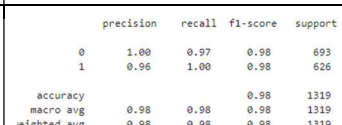
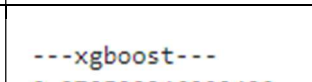
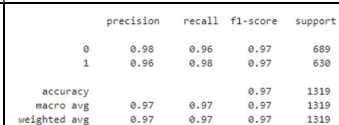
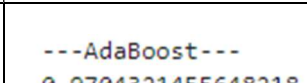
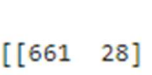
The Final Model Selection Justification articulates the rationale for choosing Gradient Boosting as the ultimate model. Its exceptional accuracy, ability to handle complexity, and successful hyperparameter tuning align with project objectives, ensuring optimal loan approval predictions.

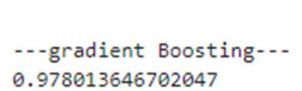
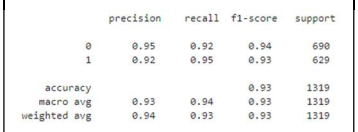
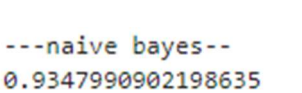
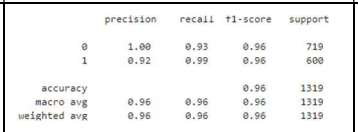
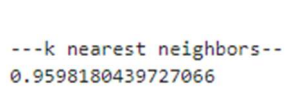
Model Optimization and Tuning Phase Report: Click Here

Results

Output Screenshots

This section includes screenshots of the model predictions, visualizations, and any relevant user interfaces.

Model	Classification Report	Accuracy	Confusion Matrix
Logistic Regression			
SVC			
Decision Tree Classifier			
Random Forest Classifier			
XGBoost			
AdaBoost			

Gradient Boost			
Naïve Bayes			
K Nearest Neighbors			

Advantages & Disadvantages:

- **Advantages:**

- Improved shipping time predictions.

- Enhanced customer satisfaction.

- Better logistics planning.

- **Disadvantages:**

- Dependency on data quality.

- Potential biases in the model.

Future Scope

The future scope for eCommerce shipment prediction using machine learning is extensive and encompasses various innovative applications and enhancements. Here are some key areas with potential for future development:

1. Enhanced Prediction Models

- **Advanced Algorithms:** Continued development and integration of more sophisticated machine learning models (e.g., deep learning, reinforcement learning) for higher accuracy.
- **Real-Time Predictions:** Utilizing streaming data to provide real-time shipment predictions and updates.

2. Integration with Emerging Technologies

- **Internet of Things (IoT):** Incorporating data from IoT devices (e.g., smart sensors on packages, vehicles) to enhance tracking and prediction accuracy.
- **Blockchain:** Ensuring transparency and traceability in the shipping process by integrating blockchain technology.

3. Personalization and Customer Experience

- **Personalized Delivery Estimates:** Providing customers with tailored delivery estimates based on their location, order history, and preferences.
- **Proactive Communication:** Offering real-time notifications and updates to customers about their shipments, improving satisfaction.

4. Optimization of Logistics and Supply Chain

- **Dynamic Route Optimization:** Continuously optimizing delivery routes based on real-time traffic, weather, and other factors to ensure timely deliveries.
- **Inventory Management:** Integrating shipment prediction with inventory management systems to balance stock levels and reduce storage costs.

5. Sustainability and Cost Efficiency

- **Eco-Friendly Shipping:** Optimizing routes and delivery methods to minimize carbon footprints and promote sustainable practices.
- **Cost Reduction:** Identifying cost-saving opportunities in the supply chain through predictive analytics.

6. Global and Cross-Border Shipping

- **International Logistics:** Enhancing predictions for cross-border shipments by accounting for customs procedures, international regulations, and multi-modal transport logistics.
- **Global Supply Chain Resilience:** Building more resilient supply chains by predicting and mitigating the impact of global disruptions (e.g., political events, natural disasters).

7. Risk Management and Mitigation

- **Predictive Maintenance:** Forecasting maintenance needs for delivery vehicles and infrastructure to prevent breakdowns and delays.
- **Disruption Prediction:** Anticipating potential disruptions in the supply chain and developing contingency plans.

8. Data Utilization and Integration

- **Big Data Analytics:** Leveraging vast amounts of data from multiple sources to improve prediction models.
- **Unified Data Platforms:** Creating integrated platforms that combine data from eCommerce, logistics, and external sources for comprehensive analysis.

9. Automation and Autonomous Systems

- **Autonomous Vehicles and Drones:** Integrating predictions with autonomous delivery systems to enhance efficiency and reduce human intervention.
- **Robotic Process Automation (RPA):** Automating routine logistics and administrative tasks using RPA and AI.

10. Regulatory Compliance and Ethical AI

- **Automated Compliance:** Ensuring shipments comply with international and local regulations through automated checks and balances.
- **Ethical AI Practices:** Implementing ethical AI guidelines to ensure fairness, transparency, and accountability in shipment predictions.

11. AI and Machine Learning Integration

- **Hybrid AI Models:** Combining machine learning with other AI techniques like natural language processing and computer vision for comprehensive shipment management.
- **Self-Learning Systems:** Developing self-learning systems that continuously improve based on new data and feedback.

Appendix

10.1. GitHub & Project Demo Link

For source codes and project file submission check the below mentioned links for the files and the links.

GitHub Repository Link: Click Here

Presentation Link: Link-1
Link-2