SMARTBRIDGE
Let's Bridge the Gap
a Veranda Enterprise

Smart Internz

Final Project Report Template

1. Introduction

1.1. Project overview

Problem Statement:
The e-commerce industry faces a common challenge: accurately predicting shipping times. Delays in delivery can impact customer satisfaction and operational efficiency. Our goal was to build a predictive model that estimates whether a product will reach its destination on time.
Our e-commerce shipping prediction model bridges data science and logistics, improving customer experience. By leveraging machine learning, we empower businesses to make informed decisions and streamline shipping processes.

1.2. Objectives

1. **Accurate Delivery Time Estimation:**
    - Develop a predictive model that estimates delivery times for e-commerce shipments.
    - Minimize delays by providing reliable estimates to customers.
2. **Enhanced Customer Satisfaction:**
    - Improve the overall customer experience by reducing uncertainty around delivery dates.
    - Enhance trust and loyalty among e-commerce users.
3. **Operational Efficiency:**
    - Optimize logistics operations by streamlining shipping processes.
    - Reduce unnecessary back-and-forth between hospitals, clinics, and pathology labs.
4. **Integration with E-commerce Platforms:**
    - Seamlessly integrate the predictive model into popular e-commerce platforms.
    - Enable real-time predictions during the checkout process.
5. **Business Impact:**
    - Quantify the impact of accurate shipping predictions on revenue, customer retention, and operational costs.

2.Project Initialization and Planning Phase

First major segment that is involved in the project is the “Project Initialization and Planning Phase.” During this, the goals and objectives of the project are stated and recognized the scope of a particular project and all consumers of the project. At this stage, it is identified who would be heading the project, how many people would be involved in the project, what resources would be available and a most probable time frame is set. Moreover, the deliverable creation include the risky assessment phase together with the development of the risk management plan. Improved initiation would also have contributed towards development of a grand and great ML idea that is set up systematically without ambiguity, and with strategies on the course to follow in case of the encountering of any difficulty.

1.3. Define Problem Statement

The factor that has been revealed to pose a very significant influence on customer satisfaction on e-commerce is referred to as the ‘on-time delivery’ factor. On the other side, in case of estimating the delivery time, the situation is somewhat more complicated due to the number of factors like weather conditions, traffic and other complexities of the logistic to which there are many factors that cannot be easily predicted accurately. Seeing this challenge, the proposed project is creation of a reliable machine learning based prediction system of delivery time in the logistics of e-commerce. The techniques such as the information system combined with links to e-commerce sites history search & update services will go a long way to assisting the customer in arriving at the correct & standard delivery estimates. Now here what we are going to do is going to add more to the entire customer experience by reducing the sources of uncertainty associated with delivery, and then enhance the logistic operation to the scalability and efficiency themes through the studies of machine learning models.

Define Problem Statement : Click Here

1.4. Project Proposal (Proposed Solution)

The outlined dependent variable would be the production of a probability model of timely delivery of products. The model will consider various factors, including. Aspects that will be catered for by the model include the following:

Shipping Methods: The methods of shipping or delivery are quite common, namely, standard and express that indicate certain delivery time.

Key Steps:

Data Collection and Preparation:

Stock information relating the shipping information and the previous products in an accurate manner.
Prepare the data set, focus on missing values that might be present in set and new features that might occur after collecting a huge amount of data.

Feature Engineering:

Create relevant features
Analyse records of the approved shipping of some of the shipments in the past in order to determine some of the trends
.

Model Selection and Training:

Broadly get aware of the existence of the ML algorithms like polynomial regression, random forest etc.
Therefore, train the model using the developed dataset.

Model Evaluation and Deployment:

Thus, it is recommended to select a common base for establishing effectiveness of the models with the help of some statistical measures such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE).
Apply the above model as a clean and simple web service to type in the new data and get the predictions in return in real time

Project Proposal : Click Here

1.5. Initial Project Planning

We shall develop a machine learning model to predict e-commerce order shipment times for improved customer service and business efficiency. This will involve defining data needs, cleaning and preparing data, selecting and training machine learning models, and optionally creating a user interface to display predictions regarding it. If done properly, it should produce more precise forecasts of deliveries to customers and an optimized allocation of resources for an enterprise. We will develop the details with stakeholders, get our data in order, and choose a machine learning method.

Initial Project Planning : Click Here

2. Data Collection and Preprocessing Phase

Data Collection and Preprocessing In this phase, a plan will be run to collect only relevant data sets for E-commerce Shipping Prediction at Kaggle. Afterwards, the quality of data is checked through data verification, and inconsistencies and missing values are treated. Preprocessing involves mainly cleaning, encoding, and organizing the dataset properly for further exploratory analysis and development of an analytical model based on machine learning algorithms.

2.1. Data Collection Plan and Raw Data Sources Identified

The dataset for " E-commerce Shipping Prediction " is sourced from Kaggle. It includes customer ratings, price of product and important information like mode of shipment. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling

Data Collection Plan and Raw Data Sources Identified : Click Here

1.1. Data Quality Report

The dataset for " E-commerce Shipping Prediction " is sourced from Kaggle. It includes customer id's and information about orders. Data quality is ensured through thorough verification, addressing missing values, and maintaining adherence to ethical guidelines, establishing a reliable foundation for predictive modeling.

Data Quality Report : Click Here

1.2. Data Exploration and Preprocessing

Data exploration and preprocessing are the initial steps in getting valuable insights from a dataset. They work together to prepare the data for further analysis, like training a machine learning model.

Data Exploration (Exploratory Data Analysis - EDA)

- **Understanding the data:** This involves identifying what variables it contains, and their data types.
- **Finding patterns and trends:** Statistical techniques and visualizations like histograms, scatter plots, and boxplots to uncover relationships between variables and identify any interesting patterns.
- **Identifying data quality issues:** This might involve checking for missing values, outliers (extreme data points), or inconsistencies in formatting.

Data Preprocessing

- **Cleaning the data:** This addresses the issues discovered in EDA. Like filling in missing values, removing outliers and fixing formatting errors.
- **Data transformation:** Transforming the data to make it more suitable for analysis. This could involve scaling numerical features, encoding categorical features, or creating new features based on existing ones.

Data Exploration and Preprocessing : Click Here

3. Model Development Phase

The Model Development Phase is where the best model that will suit the problem and data has to be chosen, followed by training on our prepared data. The model will learn on part of the data provided. During this phase, the model will learn patterns and relationships. To measure its performance and understand where it goes wrong, it will be evaluated on another set of data. This process of training and evaluation can be iterative. Among the evaluated accuracies of such a variety of models, the best is considered.

1.1. Feature Selection Report

In the initial model development, all the features available in the Ecommerce shipment details dataset were used with the exception of customer id. The reason behind this approach was to capture potential information. Ecommerce shipments involve various factors like customer ratings and previously purchased history, customer calls, etc. which individually may not directly affect the target variable but they may indirectly affect shipment time, and including all features ensured that we didn't miss any potentially important contributors to the model's goal.

Feature Selection Report : Click Here

1.1. Model Selection Report

In the evaluation process, we found Gradient Boosting to be a very promising model in predicting whether a shipment arrived on time or not, apart from performing well on the validation set. It improved the limitations that models KNN, Decision Trees, and Random Forest showed, doing better in accuracy according to the chosen metric. There are more advanced ways of tuning Gradient Boosting for better hyperparameters, which may lead to better performance.

Model Selection Report : Click Here

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

We began modeling, using a very powerful technique called Gradient Boosting. It learns how to predict from information obtained in credit card shipment data. Then, we created a separate test to make sure that the model works fine by checking the accuracy of predictions for on-time deliveries. This was checked using different factors like accuracy, precision, etc.

Initial Model Training Code, Model Validation and Evaluation Report : Click Here

4. Model Optimization and Tuning Phase

Model Optimization and Tuning is the stage at which machine learning models are tuned for improved performance. This includes optimized model code, fine-tuning hyperparameters, comparing performance metrics of models, and justifying the final model selection to ensure improved predictive accuracy and efficiency.

Now, we will develop an initial model based on gradient boosting and tune these models to achieve optimal performance. Here is how:

1. Gradient boosting has settings like the number of training rounds and learning rate. These settings could very well be thought of as dials which we can crank around arbitrarily to change exactly how our model learns from the data. We will run through different combinations of these settings to see how they affect the model's accuracy at the prediction of on-time deliveries.

2. This means the objective of model performance is neither so simple that it can get too easiness nor too complex that it can suffer overfitting. Underfitting is missing important patterns in data, while overfitting is a situation whereby the model simply memorizes examples in the training data too well and may probably fail on most unseen data.

3. In essence, we will use most of the measures applied in the validation phase, such as accuracy and precision, which let us know how well the model works with different settings. It then becomes a loop: change the settings, train the model, evaluate, and repeat until we get the combination that gives the best results.By optimizing and tuning the model, we hope to achieve:

- **Improved accuracy:** We want the model to be more precise in predicting on-time deliveries.
- **Reduced complexity:** We want a model that's powerful but not overly complex, making it easier to understand and use.

This fine-tuning process helps us get the most out of the initial model and ensure it delivers reliable predictions for our e-commerce shipment deliveries

1.1. 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

Hyperparameter Tuning Documentation : Click Here

1.1. 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.

Performance Metrics Comparison Report : Click Here

1.1. Final Model Selection Justification

Final Model Selection Justification

Following the model development process, **Gradient Boosting** has been selected as the final model for predicting on-time deliveries in the Ecommerce shipment dataset. This choice is supported by the following key factors:

1. Strong Performance:

During the validation phase, Gradient Boosting demonstrated superior performance compared to other evaluated models (KNN, Decision Tree, Random Forest) on metrics like accuracy and precision. This indicates its effectiveness in accurately predicting on-time deliveries.

2. Handling Complexity:

Gradient Boosting's ensemble nature allows it to handle complex relationships within the data compared to simpler models like Decision Trees. This is crucial for capturing the nuances that might influence on-time deliveries in the e-commerce setting.

3. Potential for Interpretability:

While Gradient Boosting models can be complex, techniques like feature importance analysis can be used to understand which factors contribute most to the model's predictions. This interpretability can be valuable for gaining insights into the drivers of on-time deliveries.

In conclusion, Gradient Boosting's combination of strong performance, ability to handle complexity, potential for interpretability, and successful optimization makes it the most suitable model for our goal of predicting on-time deliveries in the Ecommerce shipment dataset.

5. Results

Output Screenshots

Advantages & Disadvantages

Advantages of Using Machine Learning for Ecommerce shipment prediction

Better Customer Experience: Accurate shipping predictions set practical expectations with customers, avoiding the chances of delayed shipments and bringing increased customer satisfaction.

Increased Efficiency: With shipping time prediction, one can work out the best logistics and supply chain operations to reduce costs, consequently bringing overall efficiency into e-commerce businesses.

Better Inventory Management: Shipping predictions enable a business to manage its inventory better to stock the right products when required.

Reduced Shipping Costs: Optimization of routes and modes allows firms to bring down their shipping costs and offer the same advantage to customers.

Improved Visibility of Supply Chains: By providing real-time shipping predictions, firms can see the supply chain to know where the problem areas are and take necessary measures to improve them.

Competitive Advantage: A business that accurately predicts the time of delivery has a greater competitive edge than that which does not, thus attaining greater market shares and revenues.

Improved Communication: Shipping predictions facilitate improved communication with your customers by keeping them updated and knowledgeable about their shipments.

Disadvantages of Using Machine Learning for Ecommerce Shipping Prediction

Complexity: Building an accurate model of shipping predictions is a rather complex task that requires much resources and expertise.

Poor Data Quality: Inadequate data or incomplete data in the case of a shipment may result in bad shipping predictions. Such predictions may influence customer satisfaction and hamper business operations.

Unforeseen Circumstances: Thankfully, there are unforeseen circumstances— weather conditions, traffic congestion, or even carrier disruptions, which may affect shipping predictions and render them inaccurate.

Low Granularity: Shipping predictions can lack granularity in the sense that it cannot provide specifics of shipment details in terms of time or even location.

Lack of Carrier Data: As shipping prediction is made on the carrier's data, incomplete or inaccurate data may give way to poor predictions.

Over-reliance on Technology: It can result in too little human supervision or intervention, poor shipping predictions, and eventually dissatisfy customers.

Conclusion

It documented how a machine learning model developed for the prediction of on-time deliveries in an Ecommerce shipment dataset. Gradient Boosting was identified as bestnič since it is a very performing technique, able to handle complex relationships and, at the same time, possibly interpretable. Then, it completed fine-tuning the model to be effective on this task by hyperparameter tuning.

There are several advantages to the use of machine learning in e-commerce shipment prediction, such as improved accuracy, dynamic prediction, proactive exception handling, data-driven insight, and scalability. With these several advantages, a business can increase customer satisfaction, minimize chaotic situations in connection with logistics operations, and create advantages against competitors in the e-commerce marketplace.

This is only the tip of the iceberg. The model could improve through additional data sources or be further developed using more advanced machine learning methods. Also, monitoring and evaluation will be essential in continued effectiveness as changes occur within the underlying data and business environment. Altogether, machine learning is an incredibly powerful platform that e-commerce businesses can leverage to drive efficiency and success in shipment operations.

Future Scope

A project that considered historic shipping data, real-time carrier updates, and external factors; it just did not take into account any irrelevant variable like details of the products. The target of the project was areas that have a reliable record of the data including North America and Europe with standard ways of shipment using major carriers excluding niche ways and minor carriers. The current project takes into consideration weather, traffic, and public holidays, and known delays, but excludes infrequent one-time events. Technologies that will be used are technologies for machine learning algorithms, data pre processing tools, and integration technologies. No advanced AI unrelated to predictions is applied.

Appendix

1.1 Source Code : Click Here

1.2 GitHub & Project Demo Link

GitHub : Click Here

Project Demo Link : https://youtu.be/NbPdbErvJns?si=U1SzEcNqs1ugEdDc

Team Members GitHub Links :

1. SADASIVUNI KARTHIKEYAN BHARADWAJ

Link : https://github.com/Karthikeyan-Bharadwaj/Ecommerce-Shipping-Prediction-Using-Machine-Learning

2. VENKATA SUHAS BODDU

Link : https://github.com/suhas280405/Ecommerce-Shipping-Prediction-Using-Machine-Learning

3. MARELLA SHANMUKH ANANDUDU

Link : https://github.com/Shanmukh007-cell/Ecommerce-Shipping-Prediction-Using-Machine-Learning

4. SOMISETTI SRIDHAR

Link ; https://github.com/Empyrean-Empyrean/-Ecommerce-Shipping-Prediction-Using-Machine-Learning