Final Project Report Template

1. Introduction

1.1. Project overview

Problem Statement: One common problem within the e-commerce sector is an accurate, on-time shipping time prediction. Delayed delivery hurts customer satisfaction and operational efficiency. The goal is to develop a predictive model that can project if any given product will deliver on time or not. Our e-commerce shipping prediction model bridges data science and logistics to provide a better experience for customers. By applying machine learning, we empower businesses to drive better decisions and smoothen the process of shipping.

1.2. Objectives

Accurate Delivery Time Estimation:

Develop a predictive model estimating delivery times for e-commerce shipments.

Minimize delays by providing customers with accurate delivery time estimates.

Improved Customer Satisfaction

Improve the overall customer experience by reducing uncertainty around the date of delivery.

Boost the confidence and loyalty of e-commerce users.

Operational Efficiency:

Streamline logistics operations through efficient shipping processes.

Less back-and-forth between hospitals, clinics, and pathology labs.

Integration with E-commerce Platforms:

Incorporate the predictive model into popular e-commerce platforms seamlessly.

Enable real-time prediction in the checkout.

Business Impact:

Business impact on revenue, customer retention, and operational costs where shipping time predictions are accurate.

2. Project Initialization and Planning Phase

The first major segment involved in the project is the "Project Initialization and Planning Phase." This is where 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, in the deliverable creation include the risky assessment phase together with the development of the risk management plan. Improved initiation would also have contributed toward the development of a grand and great ML idea, set up systematically without ambiguity and strategies on the course to follow in the case of encountering any difficulty.

2.1. Define Problem Statement

The factors found out that make immense effects on customer satisfaction regarding e-commerce are the 'on-time delivery' factors. On the other hand, in the case of estimating the delivery time, things become a bit more complicated due to the number of factors such as weather conditions, traffic, and other complexities of logistics to which there are many factors that cannot easily be predicted with accuracy.

Seeing this challenge, the proposed project is
Creation of a reliable machine-learning-based prediction system for delivery time in e-commerce logistics. The information system, combined with links to e-commerce sites and history search and update services, will definitely help customers in arriving at the proper standard delivery estimates. Now, what we're going to do here is going to add more to the whole customer experience by reducing sources of uncertainty associated with delivery and enhance logistic operations in terms of scalability and efficiency themes through the studies of machine learning models.

Problem Statement Report:[Click Here]()

2.2. 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 the model will cater to include, among others:
Various methods for shipping or delivery are commonplace, namely standard and express indicating a certain delivery time.

The key steps:

Data Collection and Preparation:

Store stock information with relevance to the shipping information and previous products accurately.
Prepare the dataset, focus on the missing values that might be present in set and new features that might occur after collecting a huge amount of data.

Feature Engineering:

Relevant features creation.
Examine past records of approved shipping of some of the shipments to get an idea of 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 choose one common base while measuring the effectiveness of the models with some statistical measures such as Mean Absolute Error or Root Mean Squared Error.
Apply above model as clean and simple web service to type in new data and get the predictions in return in real time.

Project Proposal: [Click Here]()

2.3. 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.In problems of disease classification like this one, simply comparing the accuracy, that is, the ratio of correct predictions to total predictions is not enough. This is because depending on the context like severity of disease, sometimes it is more important that an algorithm does not wrongly predict a disease as a non-disease, while predicting a healthy person as diseased will attract a comparatively less severe penalty.

Initial Project Planning : [Click Here]

3. Data Collection and Preprocessing Phase

Data Collection and Preprocessing This is the phase where a program run will be executed in Kaggle to collect only relevant data sets related to E-commerce Shipping Prediction. After which, the quality of data is checked through data verification and treatment of inconsistencies and missing values. Preprocessing primarily involves cleaning, encoding, and organizing the
Proper dataset conditioning for further exploratory analysis and the development of an analytical model based on a machine learning algorithm.The Data Collection and Preprocessing phase is the execution of a plan to gather relevant Liver Patient data from Kaggle while ensuring high data quality by verification and handling missing values. Activities under preprocessing will include cleaning, encoding, and structuring the dataset in preparation for the next phases of exploratory analysis and model development.

3.1. Data Collection Plan and Raw Data Sources Identified

The dataset used for the problem " E-commerce Shipping Prediction " is taken from Kaggle. It contains reviews by customers, price of the product, and other details like modes of shipment. Data quality is assured through detailed verification, handling missing values, and compliance with ethical rules, so that a reliable base for predictive modeling is established.

Data Collection Plan and Raw Data Sources Identified : Click Here

3.2. Data Quality Report

The dataset used in " E-commerce Shipping Prediction " is obtained from Kaggle. It has customer ids and details of orders placed. Good data quality is guaranteed with proper checking, handling of missing values, and adherence to ethical guidelines that set a very good basis for predictive modeling.

Data Quality Report : Click Here

3.3. Data Exploration and Preprocessing

Data Exploration and preprocessing are the first steps toward gaining insight into a dataset. They collaborate to create an input that is suitable for further analysis, such as training a machine learning model.

Exploratory Data Analysis (EDA)

It involves knowing what variables the dataset contains and what data types they are.

Trends and patterns: Data statistical techniques, including visualizations such as histograms, scatter plots, and boxplots, relating variables, and using them to establish trends of interest.

Data quality issues identification: It could be checking for missing values, outliers, and inconsistencies in formatting.

Data Preprocessing

Cleaning the data: It encapsulates all the problems identified in EDA. Fill in missing values, remove outliers, and fix formatting errors.

Data transformation: We will often need to transform the data into forms better analysed. This might involve scaling numerical features, encoding categorical features, or creating new features from existing ones.

Data Exploration and Preprocessing : Click Here

4. Model Development Phase

Model Development Phase: The best model that can suit the problem and data has to be chosen and trained on our prepared data. It will learn on part of the provided data. In this phase, the model will learn the patterns and relationships, and to know how well it is performing and where it goes wrong, it has to be evaluated on another set of data. The process of training and evaluation can be iterated. Among such a variety of evaluated accuracies, the best is considered.

4.1. Feature Selection Report

Model Development Phase: The best model that can suit the problem and data has to be chosen and trained on our prepared data. It will learn on part of the provided data. In this phase, the model will learn the patterns and relationships, and to know how well it is performing and where it goes wrong, it has to be evaluated on another set of data. The process of training and evaluation can be iterated. Among such a variety of evaluated accuracies, the best is considered.

Feature Selection Report : Click Here

4.2. Model Selection Report

In the process of evaluation, we found Gradient Boosting very promising in predicting whether or not a shipment arrived on time. This includes performing well in the validation set. It improves the limitations that models KNN, Decision Trees, and Random Forest show by doing better in accuracy according to the chosen metric. There are more advanced ways to tune Gradient Boosting for better hyperparameters that may lead to better performance.

Model Selection Report : Click Here

4.3. 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 test to make sure the model works fine by checking the accuracy of the predictions of on-time delivery. This was checked using different factors like accuracy, precision, etc.

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

5. Model Optimization and Tuning Phase

Model Optimization and Tuning This is the stage where machine learning models are tuned for better performance. This might include model code optimized, tuning of hyperparameters, model performance metrics comparison, and justification of final model selection to ensure predictive accuracy and efficiency.

Now, we will develop an initial model based on Gradient Boosting and tune these models to get optimal performance. Here is how:

1. Gradient Boosting has parameters like the number of training rounds and learning rate. Literally, these could be thought of as dials that we can turn around in a pretty arbitrary fashion to change exactly how our model learns from the data. We will walk through different combinations of these settings to see the effect on the accuracy of the model in predicting on-time deliveries.
2. This means neither so simple that it gets too much easiness nor so complex that it can suffer overfitting. Underfitting is lacking important patterns in the data, while overfitting refers to a situation whereby a model simply memorizes examples in the training data too well and may probably fail on most unseen data.
3. Basically, most of the measures that will be applied at the time of validation will let us know how fine the model works under different settings. After all, it's a loop: change the settings—train the model—evaluate—repeat until we get the combination that gives the best result. Optimising and tuning the model is done to achieve:

Accuracy improved: We want a more precise model in on-time deliveries.

Less complex: We want a robust model that is not too complex; hence, being able to understand and apply it further. We wrung out our model for that last little bit to ensure good prediction in our e-commerce shipment deliveries.

5.1 Hyperparameter Tuning Documentation

Gradient Boosting was chosen for this problem because it performed better during tuning. It is capable of handling complex relationships and performing all error minimization to avoid overfitting, thus providing optimum predictive accuracy, which was in line with the objectives of the project. Thus, this ended up being the final model.

Hyperparameter Tuning Documentation : Click Here

5.2. Performance Metrics Comparison Report

The Performance Metrics Comparison Report contrasts the baseline and optimized metrics for different models, underlining the improved performance of the Gradient Boosting model. This assessment gives an exact insight into the refined predictive abilities achieved by hyperparameter tuning.

Performance Metrics Comparison Report : Click Here

5.3. Final Model Selection Justification

Final Model Selection Justification

Gradient Boosting is the final model selected to predict on-time deliveries in the ecommerce shipment dataset after going through the model development process. This was based on support from the following key factors:

1. Strong Performance:

Compared with the rest of the models that were tested during the validation phase, Gradient Boosting performed better in terms of accuracy and precision. This can be very useful for ensuring the on-time delivery of an accurate prediction.

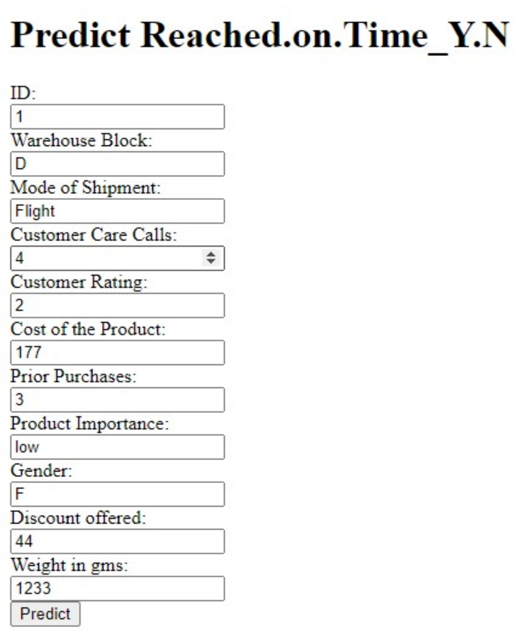
2. Handling Complexity:
The Gradient Boosting algorithm's ability to handle complex relationships in data is derived from its ensemble nature, unlike the simpler Decision Trees model. This is very important in capturing subtleties that might bear on on-time delivery in an e-commerce setting.

3. Potential for interpretability
While Gradient Boosting models may be complex, techniques such as feature importance analysis allow insight into which factors most contribute to the prediction of the model. This, in turn, is rather useful interpretability for gaining insights into what drives on-time deliveries.

In summary, good performance, handling of complexity, potential interpretability, and optimization to great success are what drive us to choose Gradient Boosting as best for our goal of on-time delivery prediction in the Ecommerce shipment dataset.

6. Result

Output Screenshots



Prediction Result

The shipment is predicted to be: **On Time**

Go back to the form

Advantages and Disadvantages

Final Model Selection Justification

Based on building the models in this study, Gradient Boosting was chosen as the final model to predict on-time deliveries within the Ecommerce shipment dataset. This choice is justified by the following key factors:

Advantages of Using Machine Learning for Ecommerce shipment prediction

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

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

Improved Inventory Management: Shipping predictions help a business maintain its inventory levels to have the right products on hand when they are needed.

Reduced Shipping Costs: Optimization of routes and modes allows firms to reduce their shipping costs and pass on this same advantage to customers.

Improved Visibility of Supply Chains: Due to the 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 correctly 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 forecasts facilitate improved communication with your customers by keeping them updated and informed about their shipments.

Disadvantages of Using Machine Learning for Ecommerce Shipping Prediction

Complexity: Shipping prediction modeling is quite a complex task that requires much resources and expertise.

Poor Data Quality: Bad shipping predictions occur due to inadequate data or incomplete data in the instance of any shipment. Such predictions have an impact on customer satisfaction and impede business operations.

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

Low Granularity: Shipping prediction lacks the granularity to provide the specifics of shipment details with respect to time or even location.

Lack of Carrier Data: Shipping prediction is done based 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 customer dissatisfaction.

Conclusion

It described the development of a machine learning model for the purpose of predicting on-time delivery using a dataset of online shipments. The technique known as gradient boosting was found to be the most effective because it can manage intricate relationships and may also be interpreted. The model was then fully adjusted by hyperparameter tweaking to make it efficient for this purpose.

The application of machine learning to e-commerce shipping prediction has a number of benefits, including increased precision, dynamic forecasting, proactive exception management, data-driven insight, and scalability. With these many benefits, a company can boost client pleasure, reduce tumultuous situations related to logistical operations, and obtain an advantage over rivals in the online retail space.

This is merely the very beginning. More data sources could help the model get better, or it could be further enhanced using more sophisticated machine learning techniques. As changes take place in the underlying data and business environment, monitoring and evaluation will also become increasingly important to maintaining effectiveness. All things considered, machine learning is a really potent tool that e-commerce companies may use to boost productivity and success in shipping operations.

Future Scope

A project factorized historic shipment statistics, current carrier updates, and outside variables; this considered discarding variables thought to be unimportant, like product information. It focused on regions that have dependable records of data, like North America and Europe. Major carriers were used for normal shipping methods which omit niche and small carriers. The project does not provide for single rare events but considers weather, traffic, public holidays, and known delays. As a part of this project, machine learning algorithms, data preprocessing tools, and integration technologies will be applied. No use of sophisticated AI unrelated to forecasts.

Appendix

1.1 Source Code : Click Here

1.2 GitHub & Project Demo Link: **Click Here**

GitHub :

Project Demo Link : **https://www.youtube.com/embed/-LL8IRaSp5w**

Team Members GitHub Links :

1) Naru Jyothika Reddy
https://github.com/jyothikan12/Ecommerce-Shipping-Prediction-Using-Machine-Learning

2) Naga Lahari Tummala
https://github.com/Laharit7/Ecommerce-Shipping-Prediction-Using-Machine-Learning

3) Satya Sai Charan
https://github.com/SaiCharan72/Ecommerce-Shipping-Prediction-Using-Machine-Learning

4) Venkata Abhishek
https://github.com/Jogendra04/E-commerce-Shipping-Prediction-using-ML