



**Name of the Project- Product Shipment Delivered on time or not ?**

# **The E-Commerce Shipping Problem**

***By Pooja Keer***

***Batch- PGA19(THANE)***

## E-COMMERCE INDUSTRY-

"Ecommerce" or "electronic commerce" is **the trading of goods and services on the internet.**

The e-commerce industry has seen significant growth in recent years, with more and more people shopping online. As a result, there is demand for efficient and reliable shipment services to deliver goods to customers in a timely manner.

One of the major challenges in e-commerce shipment is the management of the delivery process. The shipment process involves multiple stages, from receiving the order to packaging the goods and finally delivering them to the customer.

Each stage of the process must be carefully coordinated to ensure timely delivery and minimize the risk of errors or delays.

# E-COMMERCE SHIPPING PROCESS

## Workflow Diagram of E Commerce Business

This slide is 100% editable. Adapt it to your needs and capture your audience's attention.



The shipping process involves everything from receiving a customer order to preparing it for last-mile delivery. The shipping process can be broken down into three primary stages:

- Order receiving: make sure items are in stock to fulfill the order
- Order processing: verify order data and make sure it's accurate (e.g., verifying the shipping address)
- Order fulfillment: a picking list is generated and items are picked, packed, and prepared to be shipped

## E-commerce Workflow



# HOW IT AFFECTS INDUSTRY?



- **Customer dissatisfaction**: Customers expect their orders to be delivered on time. If products are not delivered on time, customers may become dissatisfied with the service and the e-commerce platform. This can lead to negative reviews and decreased customer loyalty.
- **Loss of sales**: Delayed product delivery can lead to canceled orders, which can result in a loss of sales for the e-commerce platform. Customers may also choose to shop with competitors who are better at delivering products on time.
- **Increased shipping costs**: If products are not delivered on time, the e-commerce platform may need to pay for expedited shipping or other additional costs to meet the delivery deadline. This can increase the shipping costs, which can negatively impact the company's bottom line.
- **Reputation damage**: The timely delivery of products is critical to maintaining the e-commerce platform's reputation. If products are not delivered on time, it can damage the company's reputation and lead to decreased trust from customers.

# WHAT COULD BE BENEFITS OF DELIVERING PRODUCT ON TIME ?

- **Increased customer satisfaction**: Timely delivery of products is essential for meeting customer expectations and delivering a positive shopping experience.
- **Improved brand reputation**: Consistently delivering products on time can help build a positive brand reputation for the e-commerce platform.
- **Increased efficiency and cost savings**: An efficient logistics system that ensures timely delivery can help reduce shipping and handling costs, reduce product returns and exchanges, and streamline order fulfillment processes.
- **Competitive advantage**: Timely delivery of products can be a competitive advantage for e-commerce platforms in a crowded market.
- **Improved operational performance**: Delivering products on time can help e-commerce platforms improve their operational performance.

# Solution towards the business problem

- I have worked on the dataset called e-commerce shipping Data of about 11000 of records having features as below -
- **ID** : ID Number of Customers.
- **Warehouse block** : The Company have big Warehouse which is divided in to block such as A,B,C,D,E.
- **Mode of shipment** :The Company Ships the products in multiple way such as Ship, Flight and Road.
- **Customer care calls** : The number of calls made from enquiry for enquiry of the shipment.
- **Customer rating** : The company has rated from every customer. 1 is the lowest (Worst), 5 is the highest (Best).
- **Cost of the product** : Cost of the Product in US Dollars.
- **Prior purchases** : The Number of Prior Purchase.
- **Product importance** : The company has categorized the product in the various parameter such as low, medium, high.
- **Gender** : Male and Female.
- **Discount offered** : Discount offered on that specific product.
- **Weight in gms** : It is the weight in grams.
- **Reached on time** : It is the target variable, where 1 Indicates that the product has NOT reached on time and 0 indicates it has reached on time.



The following steps are carried out –

1. Data Preprocessing
2. Data Visualization
3. Model Fitting
4. Prediction
5. Hyperparameter Tuning

## 1.Data Pre-processing-

—Imported file in CSV format by renaming the target column 'Reached.on.Time\_Y.N':'is\_late'.

—Obtained the shape of dataset i.e (10999 rows , 12 columns) by changing the all columns into lower case as it is case sensitive.

—Described the data as info below & separated the numeric & categorical columns  
Dataframe has 10999 rows and 12 columns.

1. No missing values are found.
2. There are only 2 data types, integer and object.
3. Classification target `is_late` and others we call features.



## **2.Data Visualization-**

- Data Cleansing done by checking the missing values & duplicates in dataset
- No any missing values/duplicates are found in data.
- Identification of outlier was carried out for numeric columns using boxplot & IQR (interquartile range) method where prior\_purchase column has identified outlier which is replaced by the upper & lower bound of itself.
- so , now no outlier is identified in any numeric column.

### **3. Featuring Engineering-**

—Log transformation is carried out for numeric columns in which the plots are normalized i.e. skewed to normal distribution & Numeric Values are standardized using StandardScaler

—Feature selection is carried-out by ‘CHI-SQUARE METHOD’ in which the categorical column who ‘rejects the null hypothesis’ is the important feature for further feature encoding.

—In this, the Product\_Importance column rejected the null hypothesis, thus it's the feature importance for the data & other columns 'warehouse\_block', 'mode\_of\_shipment', 'gender', 'customer\_rating' “Failed to reject the null hypothesis”.

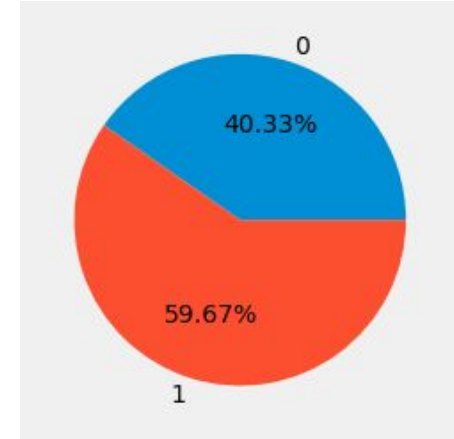
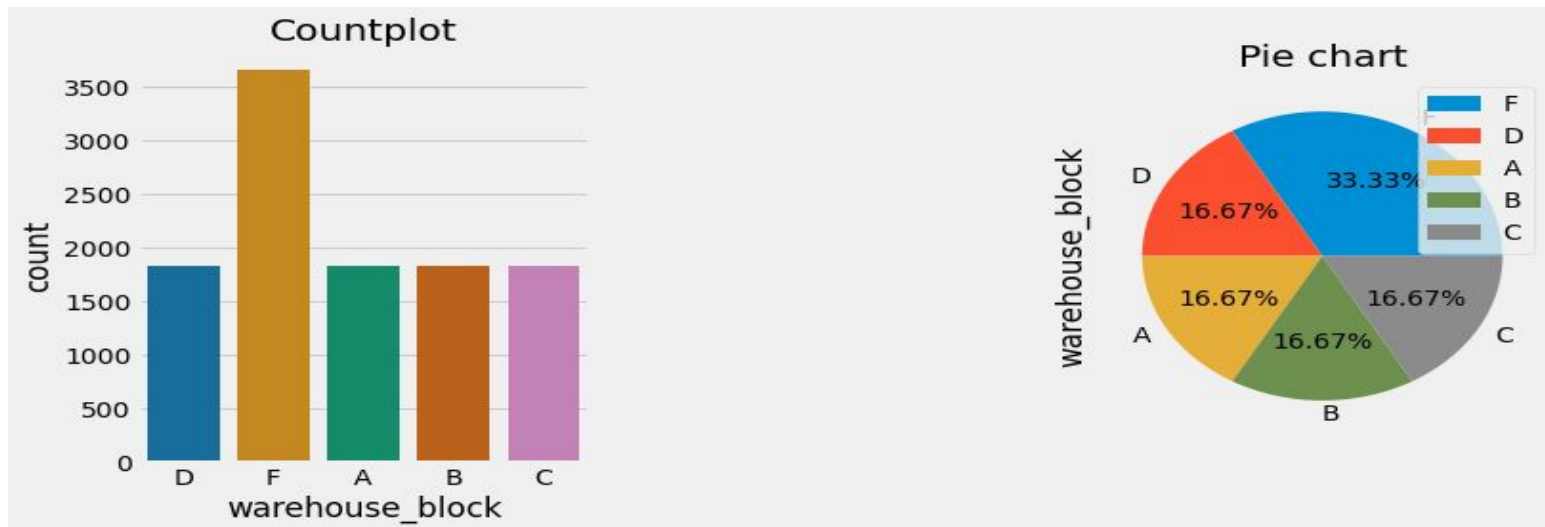
—Further, one-hot encoding is done (feature encoding) by dropping the categorical column ‘who failed to reject the null hypothesis’.

## 4.Exploratory Data Analysis-

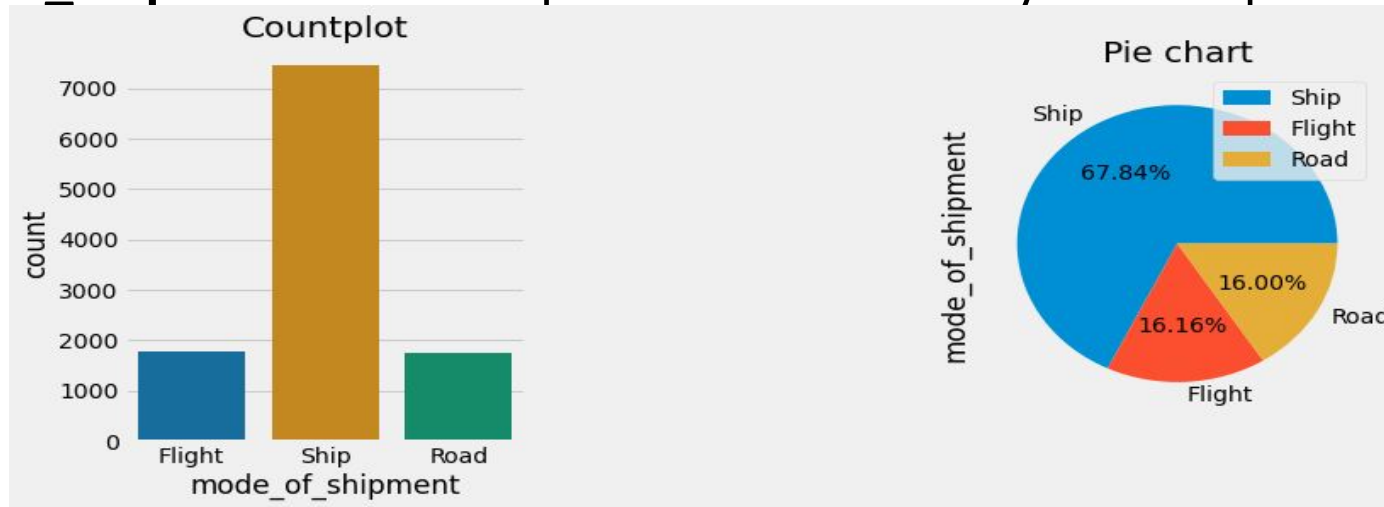
—For carrying out EDA, the copy of data is used where ‘TARGET VISUALISATION’ is done & target of class looks balanced.

—Descriptive Statistic is carried out for categorical column to get the value counts of each column & plotted the countplot as shown below-

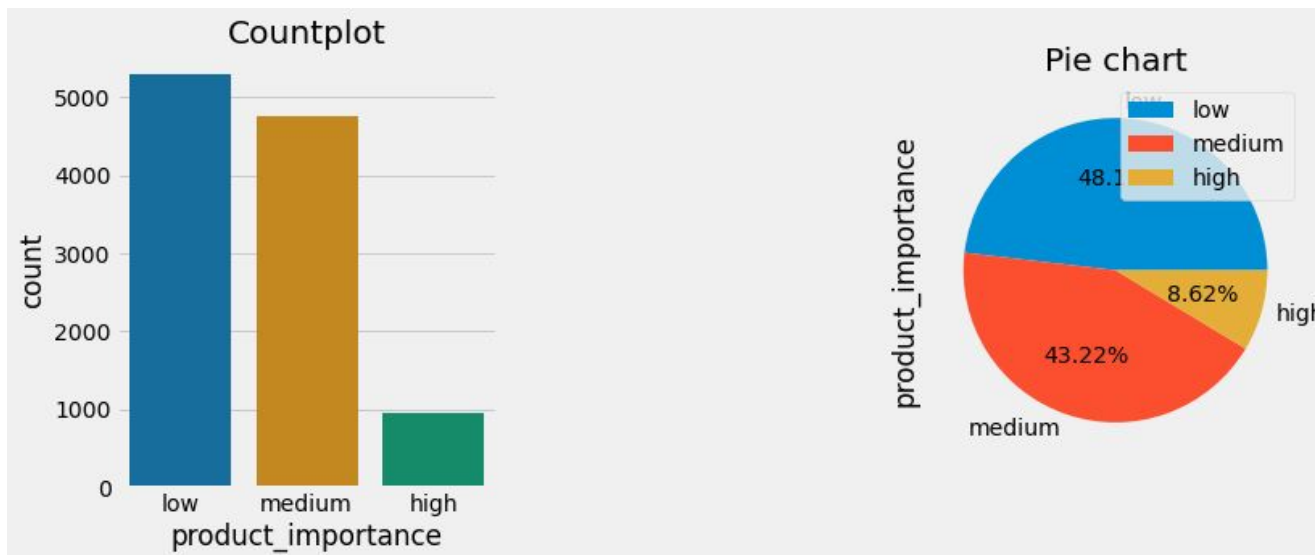
1) **Warehouse\_Block** has 5 unique values and dominated with **Warehouse\_block\_f**.



II) **Mode\_of\_Shipment** has 3 unique values and mostly used ship.



III) **Product\_importance** has 3 unique values and mostly priority of products are low.



- After doing EDA of numeric ,Distribution of **customer\_care\_calls**, **Customer\_rating**, **Cost\_of\_the\_Product**, **Prior\_purchases** is normal, because the mean and the median are close, while **discount\_offered** and **weight\_in\_grams** are indicated skewed.
- The Correlation Heatmap is obtained for studying the correlation with target variable as follows-

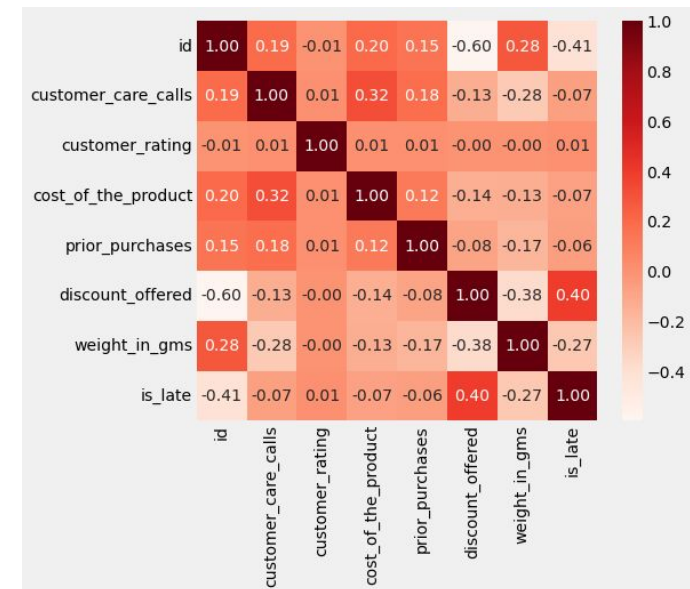
**Based on the Correlation heatmap above :**

1. Target *is\_late* has a moderate positive correlation with *discount\_offered* & weak negative correlation with *weight\_in\_gms*.
2. Feature *customer\_care\_calss* has a weak positive correlation with *cost\_of\_the\_product* and negative correlation with *weight\_in\_gms*.
3. Feature *discount\_offered* has a moderate negative correlation with *weight\_in\_gms*.

—Further, graphs were plot based on categorical columns'*mode\_of\_shipment*', '*warehouse\_block*', '*product\_importance*', '*gender*', '*customer\_rating*'-

**where we got the output-**

- Most of parcels are stored in warehouse\_block F.
- The ship contributes the most late delivery.
- Most of parcels in all shipment priority are delivered late.



## 5. Model Fitting-

–Through 'Def Fit\_evaluation ' function the model were initiated & evaluated.  
& through 'Model Comparison' function the following models were passes through dataframe -

LogisticRegression,DecisionTreeClassifier,RandomForestClassifier,  
KNeighborsClassifier,Support Vector Classifier.

with Accuracy\_Train','Accuracy\_Test','Precision','Recall\_Train','Recall\_Test',  
'F1 Score','AUC': - where the result is as-

Only **Logistic Regression and SVC** which are **neither overfitting nor underfitting**. Logistic Regression has the highest recall.

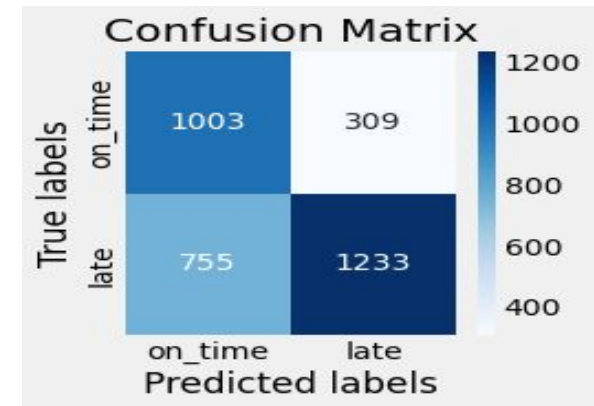
	Model	Accuracy_Train	Accuracy_Test	Precision	Recall_Train	Recall_Test	F1 Score	AUC
0	Logistic Regression	0.63	0.63	0.67	0.75	0.75	0.71	0.6
1	Decision Tree	1.0	0.66	0.72	1.0	0.71	0.71	0.64
2	Random Forest	1.0	0.67	0.76	1.0	0.66	0.71	0.67
3	KNN	0.76	0.64	0.72	0.77	0.67	0.69	0.64
4	SVC	0.68	0.66	0.88	0.52	0.51	0.65	0.7

## 6. Parameter tuning-

	Model	Accuracy_Train	Accuracy_Test	Precision	Recall_Train	Recall_Test	F1 Score	AUC
0	Logistic Regression	0.59	0.6	0.6	1.0	1.0	0.75	0.5
1	Decision Tree	0.67	0.68	0.8	0.62	0.62	0.7	0.69
2	Random Forest	1.0	0.67	0.76	1.0	0.66	0.71	0.67
3	KNN	0.81	0.65	0.72	0.82	0.69	0.7	0.64
4	SVC	0.59	0.6	0.6	1.0	1.0	0.75	0.5

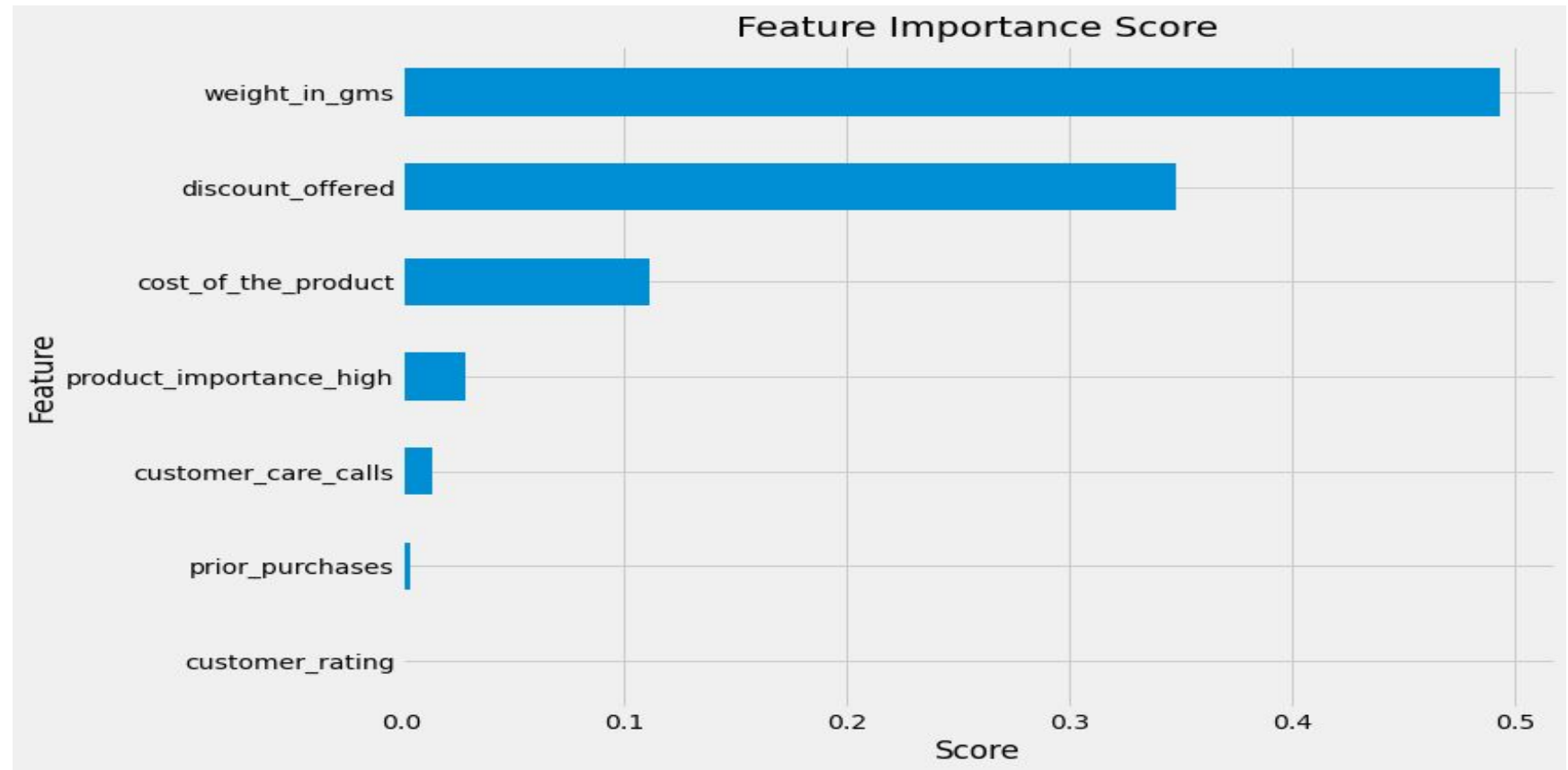
–Decision Tree algorithm with hyper-parameter tuning has a good balance between its score, also neither underfitting nor overfitting.

–Confusion Matrix after parameter tuning-





## 7.Feature Importance-



- The operation team should add more manpower when there is a sale program, especially for the discount more than 10% and the parcel weight is 1 - 4 Kg.
- The parcel should not be centralized in the warehouse block F, so that the handling is not too crowded which can cause the late shipment.
- Adding more features can improve model performance, such as delivery time estimation, delivery date, customer address, and courier.



# CONCLUSION

- The timely delivery of products can also have a significant financial impact on the e-commerce business. If products are delivered on time, it can lead to increased sales, repeat business, and revenue growth. However, if products are consistently delivered late, it can lead to canceled orders, lost sales, and increased shipping costs, which can negatively impact the company's bottom line.
- Late delivery of products can have legal consequences, particularly if there are contractual obligations regarding delivery times. Failure to meet these obligations can result in breach of contract lawsuits, financial penalties, and damage to the business reputation.
- Timely delivery of products is critical to maintaining a positive business reputation. If products are consistently delivered on time, it can lead to a positive brand reputation and increased customer trust. However, if products are frequently delivered late, it can damage the business reputation and lead to decreased customer trust.

## FURTHER ENHANCING MODEL

- The model Applied for the business problem were machine learning based. By maintaining accuracy ,further we can enhance by several deep learning models that could be useful for predicting delivery of a product on time.
- Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks,Autoencoders.
- Time-series models are useful for analyzing data that changes over time, such as the frequency and timing of deliveries.You could use time-series models to predict the delivery time based on historical data, identifying patterns and trends that can help you make accurate predictions.
- RapidMiner can be useful for predicting the delivery of products on time. RapidMiner is a data science platform that offers a variety of tools and techniques for data preprocessing, modeling, and evaluation. By leveraging RapidMiner's predictive modeling capabilities, you can build and train a model that can predict the likelihood of delivery delays based on various factors such as shipping method, location, product type, and weather conditions.



THANK YOU