BIA | BOSTON INSTITUTE OF ANALYTICS ®

E COMMERCE PRODUCT DELIVERY PREDICATION

E COMMERCE PRODUCT DELIVERY PREDICATION

INTRODUCTION

E-commerce is growing very fast
People expect quick and timely delivery.
But sometimes deliveries get delayed.

- **Late deliveries reduce customer satisfaction.**
- **This project uses machine learning to predict if a delivery will be on time or late.**
- **It helps companies plan better and improve delivery services**

BIA | BOSTON INSTITUTE OF ANALYTICS ®

E Commerce Product Delivery

INTRODUCTIONIO

Late delivery is a challenge in e-commerce.
We predict if an order will be on time.

DATA COLLECTION

Collected e-commerce delivery data.
Cleaned and prepared it for model use.

EXPLORATORY DATA ANALYSIS

Used graphs to find useful patterns.
Checked what affects delivery time

FEATURE EXTRACTION

Split data into input (X) and output (Y).Prepared features for Machine Learning model.

METHODOLOGY

Compared models to find the best one.

EXPERIMENTAL RESULTS

Model performance was compared using graphs.
Accuracy and confusion matrix were used for evaluation.

CONCLUSION

The model helps avoid delivery delays.
Can improve customer satisfaction.

APPENDICES AND REFERENCES

Dataset source and model deployment link/code

DATA COLLECTION

TOTAL ORDERS (ROWS)
10,999

TOTAL FEATURES
(COLUMNS) 12

REFINEMENT STEPS

- Warehouse Block
- Mode of Shipment
- Customer Care Calls
- Customer Rating
- Cost of the Product
- Prior Purchases
- Product Importance
- Gender
- Discount Offered
- Weight in gram

Y (Target Variable)

On-time Delivery (1 = Yes, 0 = No)

REFINEMENT STEPS

- Removed missing/null values
- Converted categorical values to numeric

  (e.g., "F" = 0, "M" = 1)
- Normalized features for better model performance

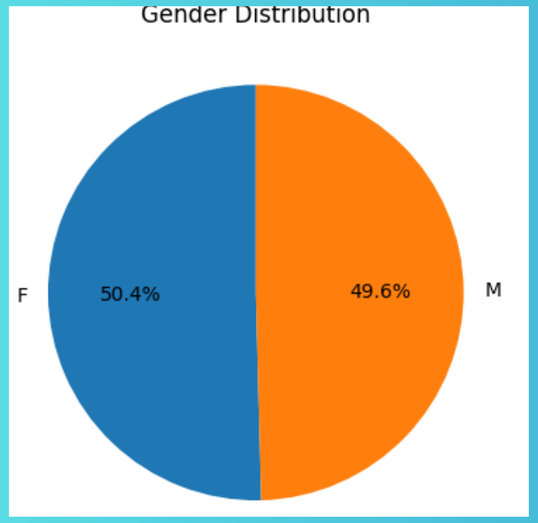
PURPOSE OF THE PROJECT:

- To predict whether an e-commerce product will be delivery on time or late.
- Help companies improve their delivery planning and customer satisfaction.

PROBLEM IT SOLVES:

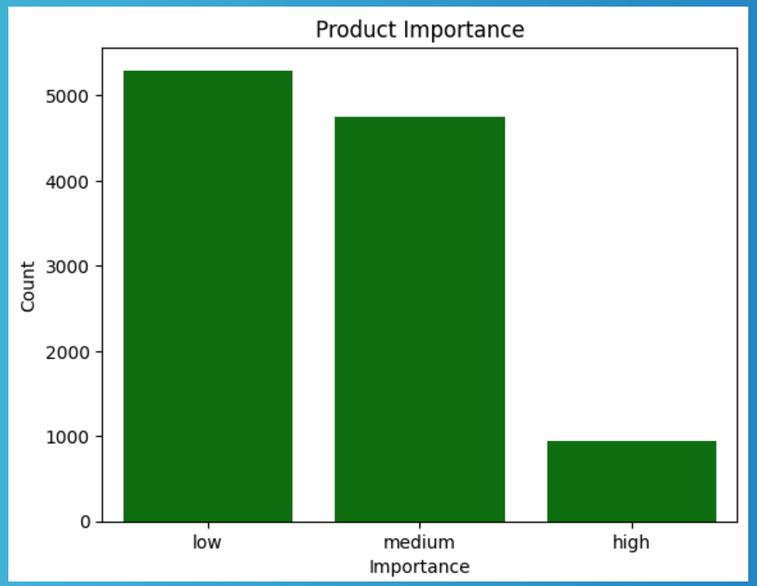
- Late deliveries cause customer complaints and cancellations.
- With this model, risky orders can be identified before dispatch, reducing delay chances.

EXPLORATORY DATA ANALYSIS

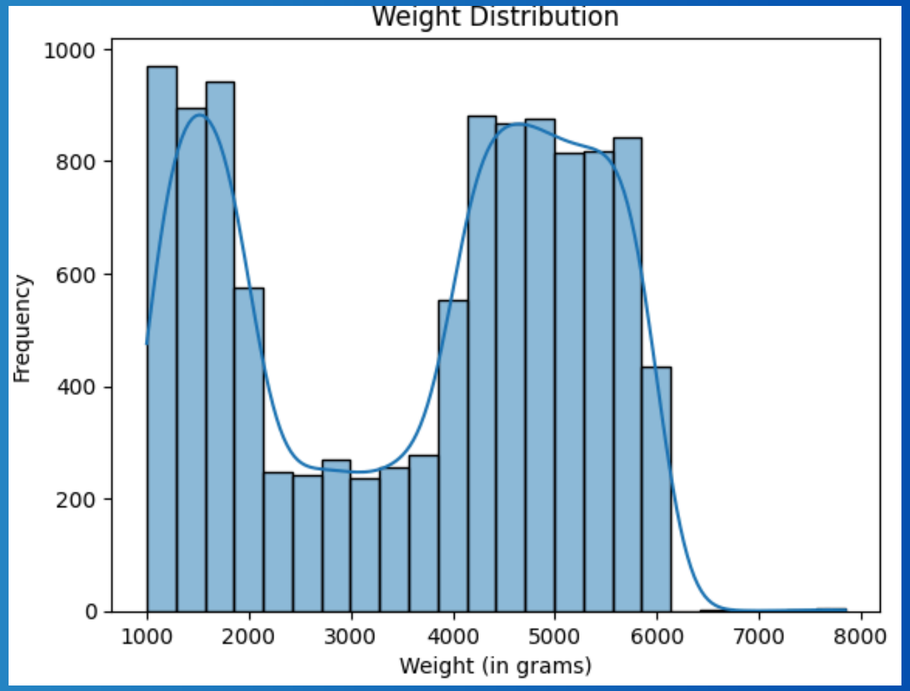


- **Female customers 54.3%, Male customers 45.7% Dataset**

shows slightly more female customers placed orders.



- **The count plot shows that medium importance products are most common. High importance products were fewer and often on-time.**



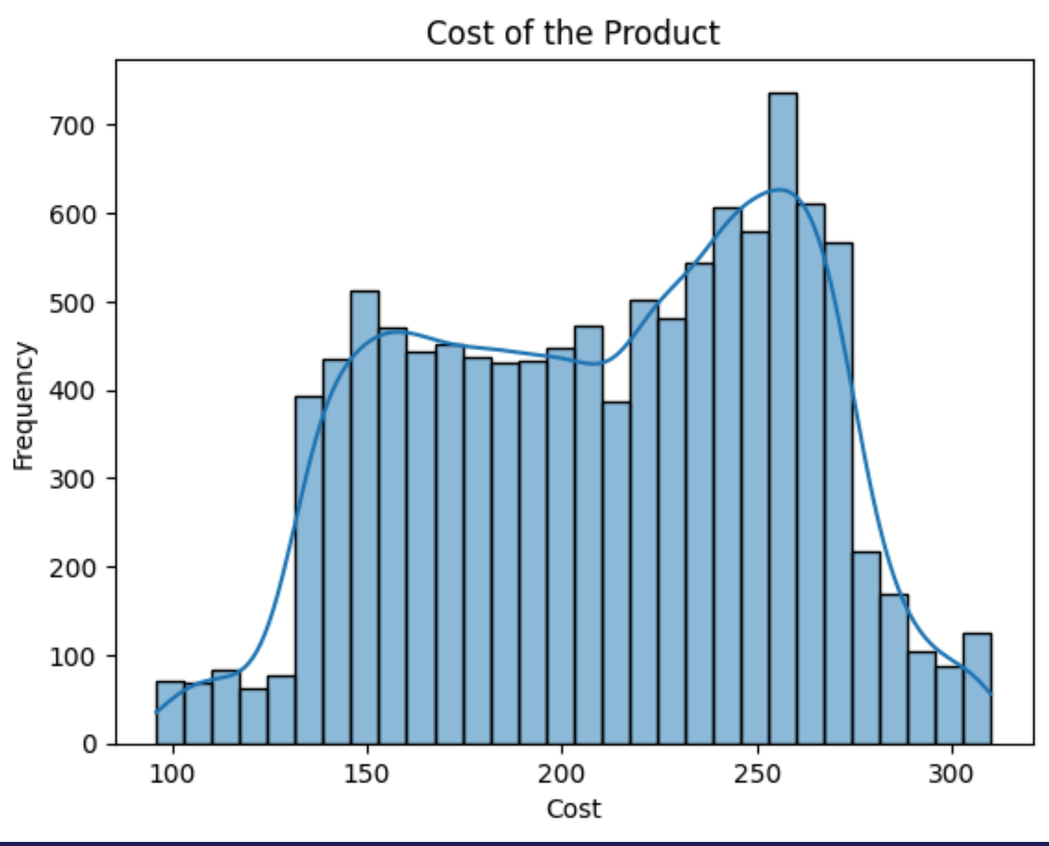
- **This chart shows how product weights are distributed across orders. Most products fall between 1500–3000 grams.**

BIA BOSTON INSTITUTE OF ANALYTICS

- The cost distribution reveals that most products are priced between ₹100 and ₹300 Very high-cost products were rare in the dataset.

- A bar chart was used to show the number of late vs on-time deliveries. Most deliveries were on time, but a significant portion experienced delays.

- Majority of the orders in the dataset were delivered on time. On-time deliveries were mostly linked to flight shipment and medium product importance.



FEATURE EXTRACTION

TRAIN - TEST AND SPLIT

- We divided the data into training (80%) and testing (20%) sets.
- random_state was set for consistent results, and stratify = y ensured balanced class distribution (on-time vs late) in both sets.

SPLINTING THE DATA INTO X & Y

- The dataset was split into two parts: X (features) and y (target variable).
- X included shipment mode, rating, cost, etc., and y represented whether the product was delivered on time.

BIA BOSTON INSTITUTE OF ANALYTICS ®

MODEL SELECTION

MODELS USED:

- **LOGISTIC REGRESSION:**

Logistic Regression is commonly used for binary classification problems, like predicting on-time or late delivery. It provides a simple and efficient way to model the relationship between independent variable and the probability of timely delivery.

- **DECISION TREE:**

Decision Tree algorithm is effective for classification tasks due to its rule-bastructure. It is easy to interpret, computationally efficient, and works well with both numerical and categorical data.

- **K-NEAREST NEIGHBORS (KNN):**

KNN is an instance-based learning algorithm that classifies a new point based on the majority class of its neighbors. It performs well on clean datasets and helps identify delivery trends based on similar past orders.

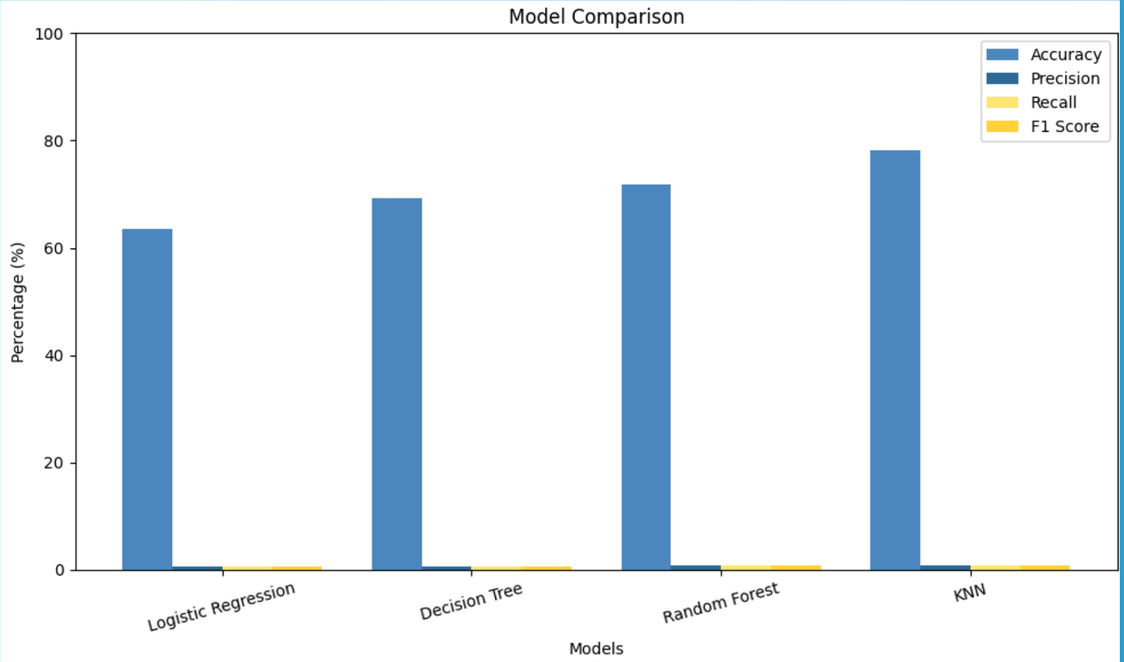
- **RANDOM FOREST ALGORITHM:**

Random Forest is an ensemble method combining multiple decision trees. It provides high accuracy, reduces overfitting, and is suitable for complex datasets like e-commerce delivery prediction.

EXPERIMENTAL RESULTS

The dataset was split into 80% training and 20% testing for evaluation.

- **The dataset was split into 80% training and 20% testing for evaluation**
- **We used four classification models to predict whether a delivery would be on time.**



Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	63.60%	0.62	0.63	0.62
Decision Tree	69.20%	0.68	0.69	0.68
Random Forest	71.90%	0.72	0.72	0.72
KNN (Final) ✅	**78.20%**	**0.78**	**0.78**	**0.78**

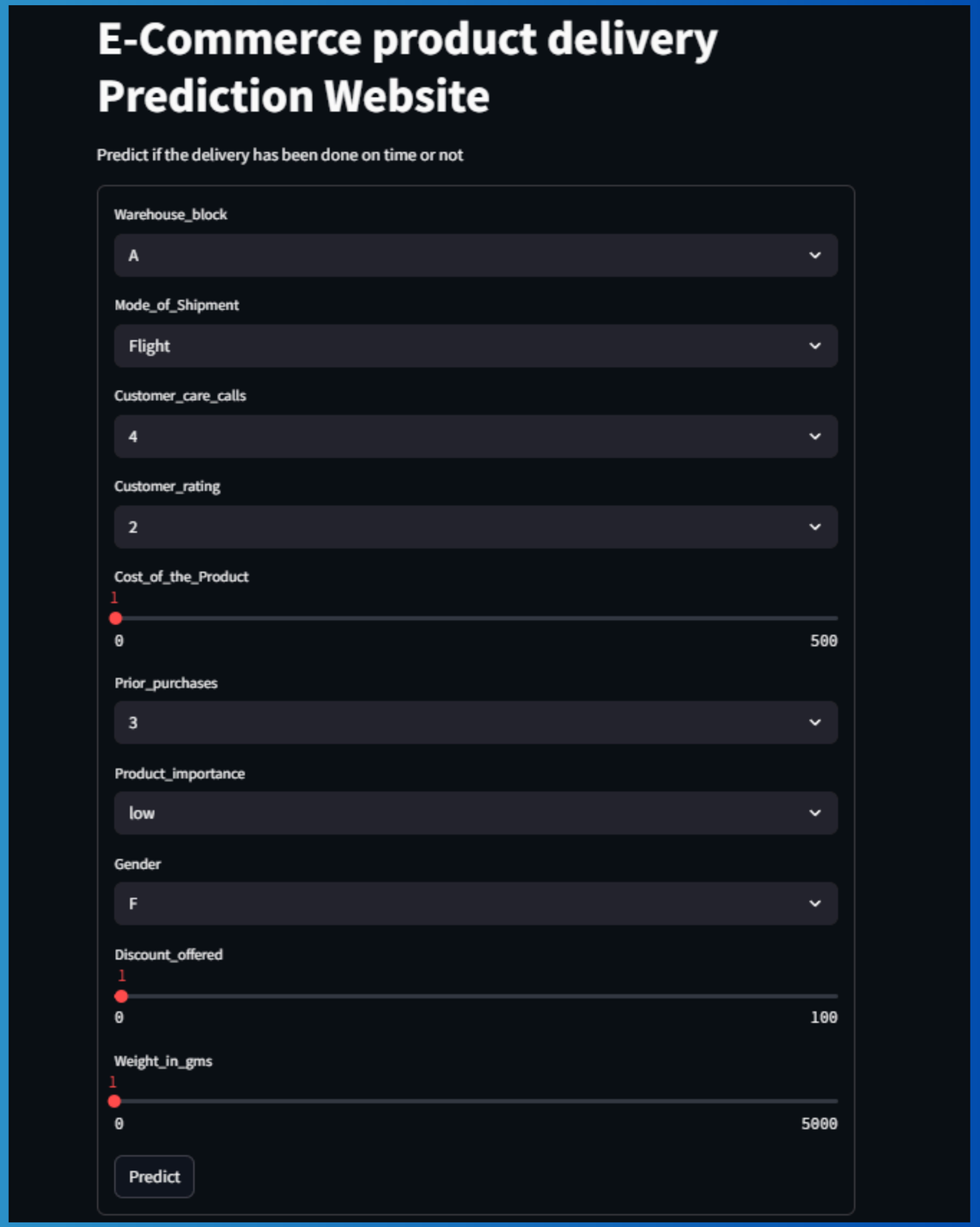
- **KNN performed the best with an accuracy of 78.2%.**
- **Therefore, K-Nearest Neighbors was selected as the final model for deployment.**

BIA BOSTON INSTITUTE OF ANALYTICS

CONCLUSION

- The project predicts whether a product will be delivered on time using machine learning.
- Among all models tested, K-Nearest Neighbors (KNN) gave the best accuracy of 78.2%.
- The model helps improve delivery planning and customer satisfaction.
- In future, we can enhance the system with real-time data and more features.

SYSTEM OUTPUT




Thank You!