Data
Science

Food Delivery Time Prediction

MUHAMMAD
UMAR
ABDURRAHMAN

Table of Content

Description

In the digital age, food delivery services are increasingly popular because they are practical and fast. The main challenge is accurate delivery time estimation. Incorrect predictions can reduce customer satisfaction. With Machine Learning, we can build models that predict delivery times based on various factors, making services more accurate and improving user experience. For example:

- Distance between restaurant and customer
- Traffic congestion
- Order time (peak/off-peak hours)
- Number of items in the order
- Courier characteristics

Project Objectives

The main objective of this project is to create an accurate food delivery time prediction model so that it can:

1. Help food delivery companies improve customer satisfaction.
2. Provide realistic estimates for customers.
3. Help optimize the company's logistics flow.
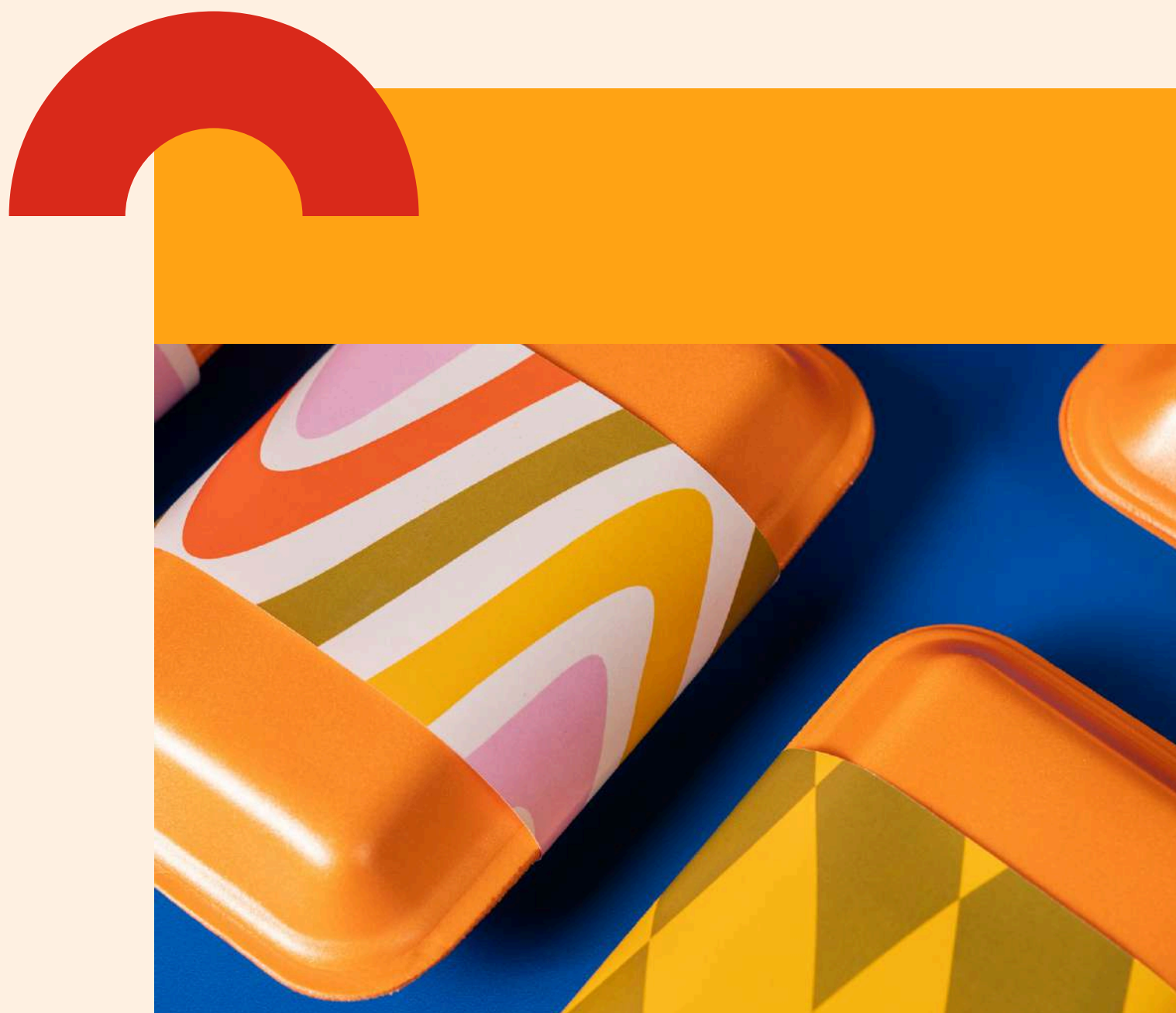
Data Understanding

Column & Row

1000 Row & 9 Column

Features

Order_ID, Distance_km, Weather, Traffic_Level, Time_of_Day, Vehicle_Type, Preparation_Time_min, Courier_Experience_yrs, Delivery_Time_min

Dataset

Food Delivery Time

Handling Missing Value → **Handling Duplicated** → **Encoding Data**

Identified minor missing values → dropped to maintain consistency.

Duplicate Check → No duplicate records found

Converted categorical features into numeric values for modeling.
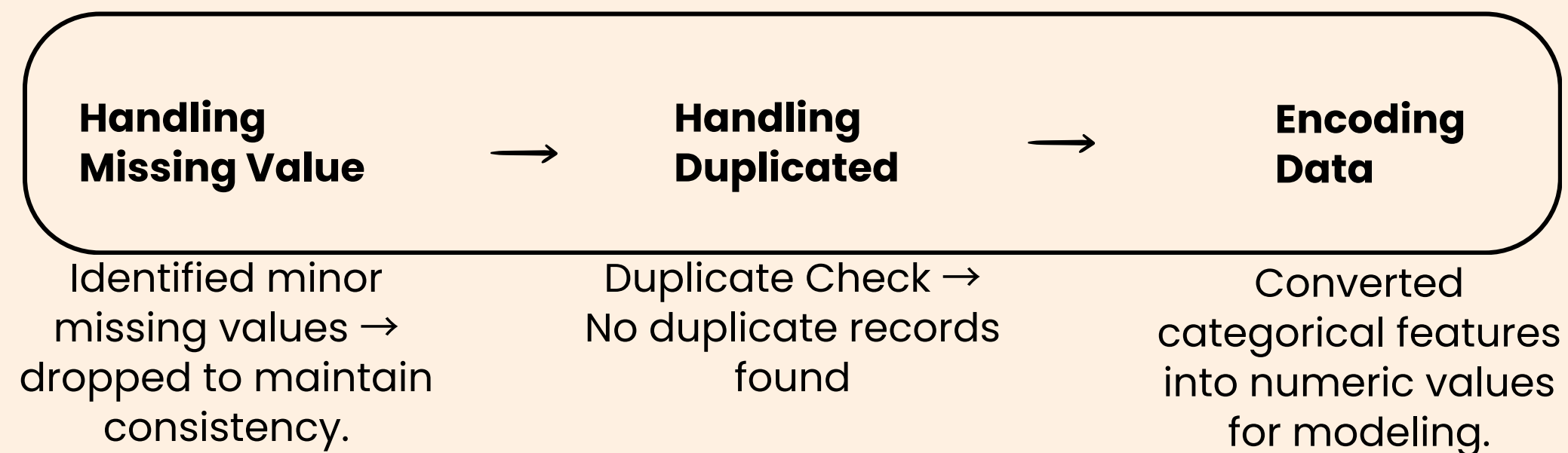
Data Pre-Processing

Distribution Delivery Time

- Distribution Shape → Right-skewed, most deliveries within 40–70 minutes, few extreme delays (>120 minutes).
- Mode → Peak at 45–55 minutes, majority of deliveries completed in this range.
- Range → From ~10–15 minutes up to 150 minutes, with potential outliers in very long deliveries.
- Spread → Most data concentrated below 100 minutes, frequency drops sharply afterward.
- Business Implication → Standard estimate can be set at 45–60 minutes; outliers (>100 minutes) require investigation (traffic, distance, restaurant prep).

Distribution Delivery Time

Correlation Heatmap Numeric Features

- **Distance (0.78)** → Strong positive correlation, the main driver of delivery time.
- **Preparation Time (0.31)** → Moderate correlation, affects delivery but less than distance.
- Courier Experience (−0.09) → Weak negative correlation, slight impact on faster deliveries.
- No Multicollinearity → Predictors are independent, each contributes unique information.

Correlation Heatmap (Numeric Features Only)
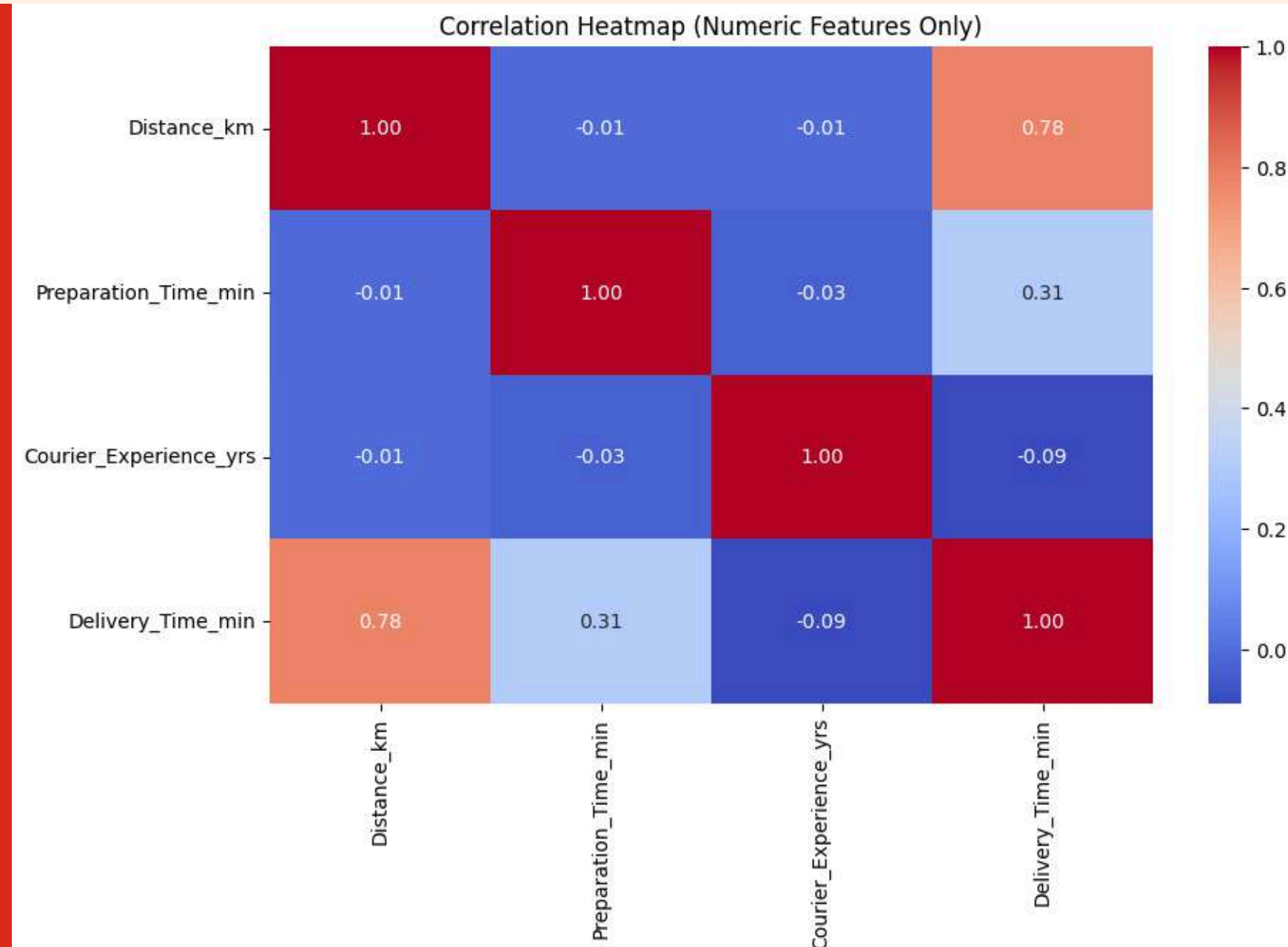
Distance vs Delivery Time

- The greater the distance → the longer the delivery time.
- Distance has been proven to be the main factor determining delivery time.
- There are variations in the same distance → influenced by other factors (weather, traffic, courier).
- There are several outliers with very long delivery times (>120 minutes).
- Distance must be included as a main variable in the prediction model.

Scatter Plot: Distance vs Delivery Time
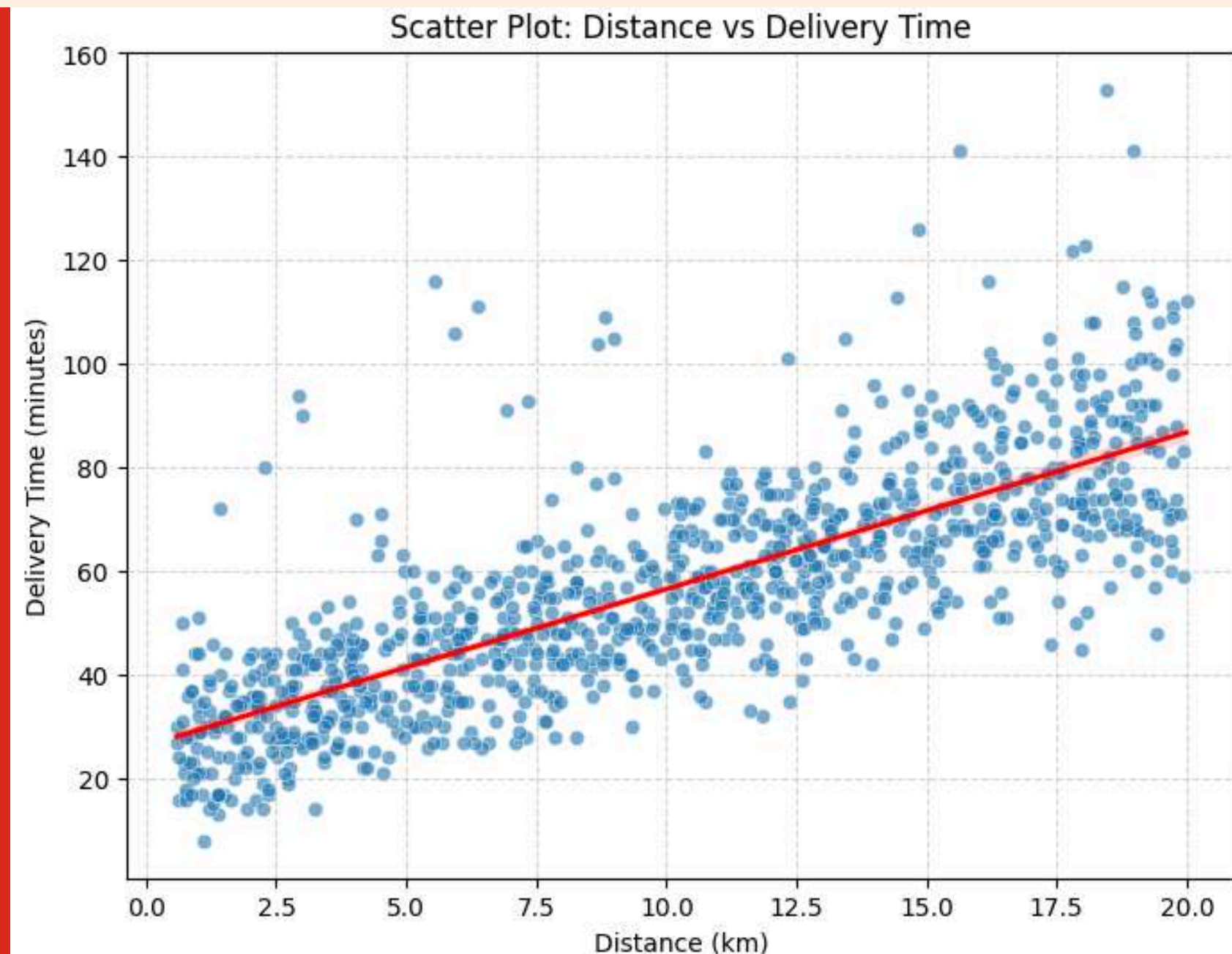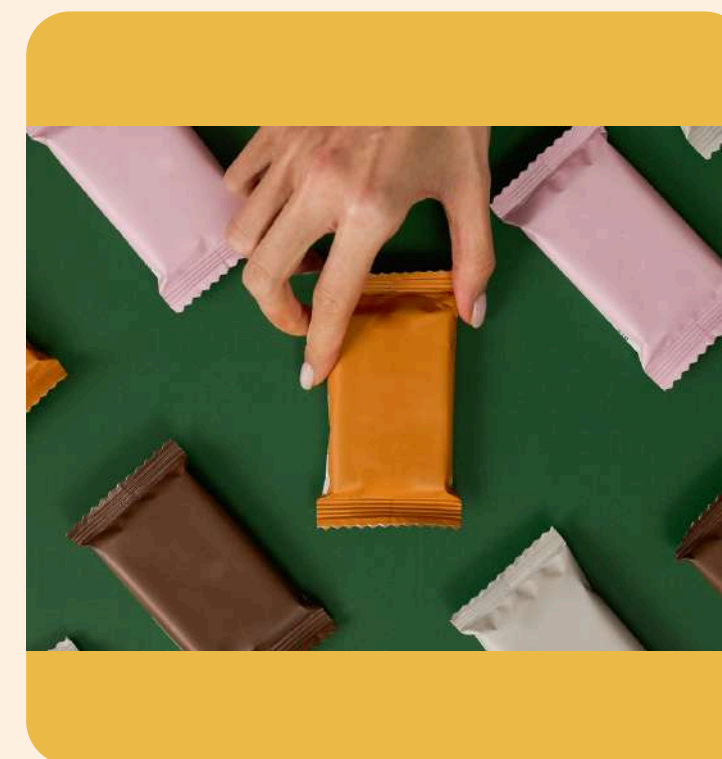
Model Preparation

Data Splitting →
- **Train (80%)**
- **Test (20%)**

Modelling Linear Regression & Random Forest

Evaluate Model Linear Regression & Random Forest

```
=== Evaluasi Linear Regression ===
MAE  : 7.29
RMSE : 10.45
R2   : 0.76
```

```
=== Evaluasi Random Forest ===
MAE  : 6.82
RMSE : 9.78
R2   : 0.79
```

Feature Importance Random Forest

- Distance (km) → Most important factor (~73%), primary driver of delivery time.
- Preparation Time (min) → Second most important (~17%), strongly affects efficiency.
- Courier Experience → Small impact (~5%), more experienced couriers slightly faster.
- Weather, Traffic, Time of Day, Vehicle Type → Very low importance (<5%), limited effect due to data imbalance.

Feature Importance - Random Forest (Tuned)
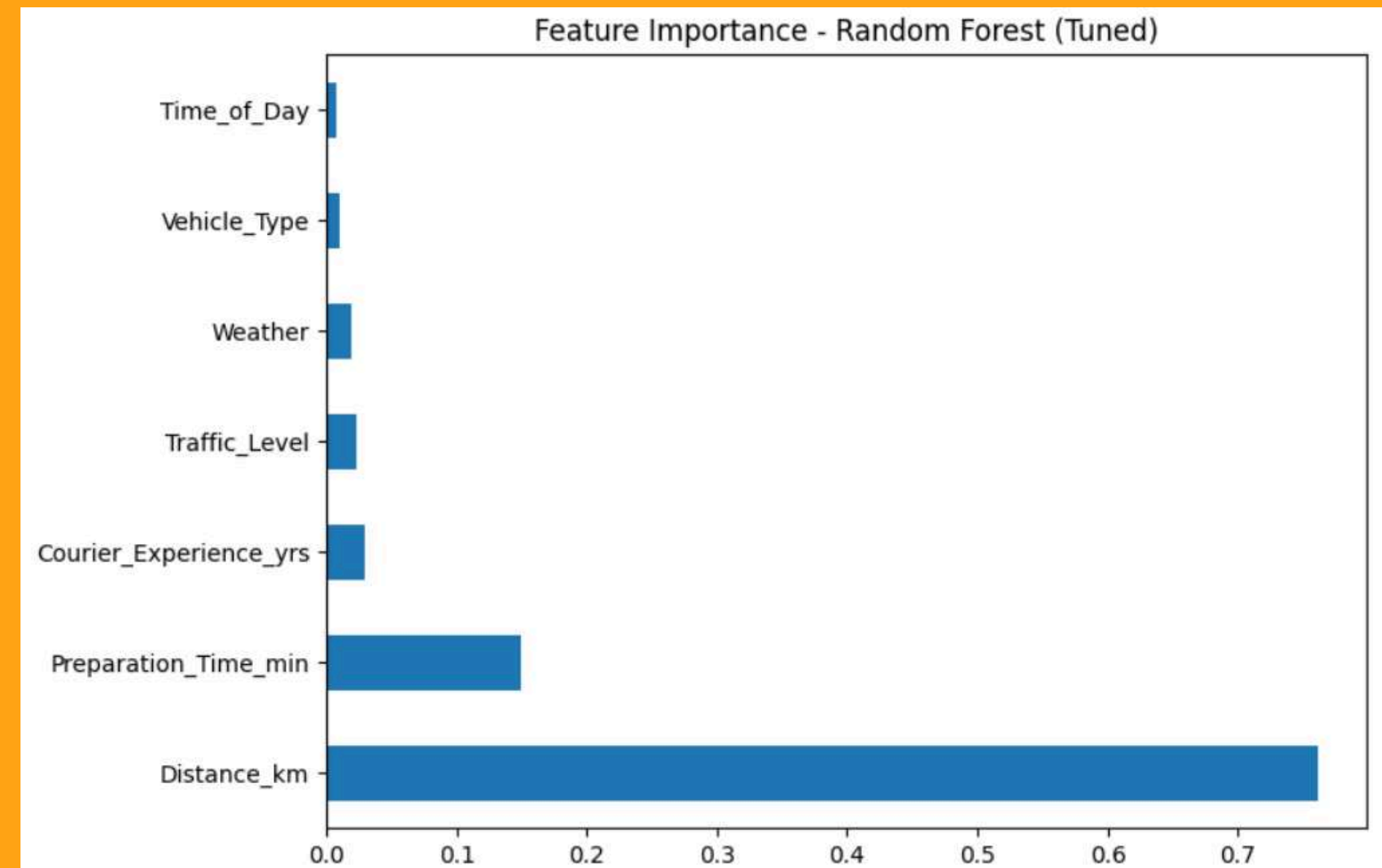
Visualization of Prediction vs Actual

Prediksi vs Actual Delivery Time

Prediction vs Actual

- Red Line (Ideal Fit) → perfect condition, prediction = actual.
- Linear Regression & Random Forest → similar performance, many points overlap and follow the red line.
- Normal Range (20–90 minutes) → fairly stable and accurate predictions, points clustered around the red line.
- Outlier (>100 minutes) → greater prediction error, model less accurate in extreme cases.

Business Recommendations High Priority

Optimize Long-Distance Orders

→ Since distance is the main driver of delays (~73% impact), consider setting dynamic delivery fees, service zones, or incentives for couriers handling long-distance deliveries.

Improve Restaurant Preparation Efficiency

→ Preparation time has ~17% influence; training staff, streamlining kitchen processes, and monitoring slow-performing restaurants can significantly cut delays.

Set Realistic Delivery Estimates

→ Use predictive model output (45–60 minutes as standard range) to provide accurate delivery time estimates, reducing customer complaints.

Business Recommendations Mid Priority

Leverage Courier Experience

→ Even though courier experience has smaller impact (~5%), assigning more experienced couriers during peak hours or long-distance deliveries can reduce outliers in delivery time.

Investigate and Manage Outliers

→ Deliveries >100 minutes often result from traffic, weather, or operational inefficiencies. Implement monitoring and alert systems to quickly address abnormal delays.

Customer Communication

→ Proactively notify customers about potential delays (traffic, weather, distance) to maintain trust and satisfaction.

Business Recommendations Low Priority

Data-Driven Logistics Planning

→ Use the prediction model to simulate scenarios (e.g., new delivery zones, vehicle types) before rollout, ensuring better operational decisions.

Conclusion

- **Built Machine Learning models (Linear Regression & Random Forest) to predict food delivery time with good accuracy (20–90 mins range).**
- **Distance (~73%) is the most influential factor, followed by Preparation Time (~17%); other factors (traffic, weather, courier experience) have smaller impact.**
- **Model provides realistic delivery time estimates (45–60 mins) → helps reduce customer dissatisfaction.**
- **Insights support operational optimization (restaurant efficiency, courier assignment, logistics planning).**
- **Demonstrates how data-driven solutions improve both customer experience and business efficiency in food delivery services.**

Thank You!
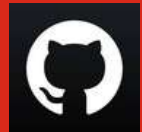
Github

muhammadumarabdurra
hman10@gmail.com

www.linkedin.com/in/u
marabdurrahman