

Amazon Delivery Prediction Report

1. Introduction

This report presents an analysis of delivery times using machine learning. The objective is to predict whether a delivery will be late (>30 minutes) based on various factors such as agent details, order time, traffic conditions, and vehicle type.

2. Data Overview

The dataset consists of multiple features related to delivery operations, including:

- **Agent Information:** Age, Rating
- **Order Details:** Order Time, Pickup Time, Category
- **External Factors:** Weather, Traffic, Area
- **Delivery Outcome:** Delivery Time (minutes)

3. Data Preprocessing

- Missing values were removed.
- Time-related features were converted to numerical formats.
- Categorical variables were encoded using Label Encoding.
- Standardization was applied to numerical variables for better model performance.

4. Exploratory Data Analysis

- The distribution of delivery times shows a peak around **120** minutes.
- The average delivery time per vehicle type is:
 - **Motorcycle: 131.03** minutes
 - **Scooter: 116.36** minutes
 - **Van: 116.14** minutes
- The average delivery time per category varies significantly. The longest delivery times are observed for **Cosmetics (132.94 minutes)**, **Kitchen (132.36 minutes)**, and **Sports (132.31 minutes)**, whereas **Grocery** has the shortest average delivery time at **26.54 minutes**. An ANOVA test resulted in a p-value of **0.0**, indicating a statistically significant difference in delivery times across categories. This suggests that product type strongly influences delivery speed, possibly due to factors such as handling requirements or logistics complexity.

5. Machine Learning Model

The **Random Forest Classifier** was used to predict delivery delays. The dataset was split into training (80%) and testing (20%) sets. The model was trained using the following features:

- Agent Age, Agent Rating
- Order Time, Pickup Time
- Weather, Traffic, Vehicle, Area, Category

Model Performance:

- **Accuracy: 99.43%**
- **Precision: 96.0%**
- **Recall: 97.0%**
- **F1-Score: 96.5%**

The confusion matrix is as follows:

```
[[ 343    17]
 [   33 8326]]
```

6. Conclusion

- The analysis shows that delivery times vary significantly depending on the product category. Categories like **Cosmetics, Kitchen, and Sports** tend to have longer delivery times, while **Grocery** is delivered much faster on average.
- The **ANOVA test (p-value: 0.0)** confirms that product category significantly affects delivery time.
- The machine learning model achieved **99.43% accuracy**, effectively predicting late deliveries based on key factors such as traffic, weather, vehicle type, and order details.
- These insights can help optimize delivery strategies, improve efficiency, and enhance customer satisfaction.

7. Next Steps

- The model provides a reliable prediction of delivery delays with an accuracy of **99.43%**.
- Further improvements can be made by including additional features such as real-time GPS tracking.
- The cleaned dataset has been exported for visualization in Power BI.

Prepared by: Mochamad Raffli Ispriyanto **Date:** 2025-03-11