Blinkit AIML Project – Project Report

This report explains the logic behind the attribution modeling used for marketing analysis and the machine learning approach used for delivery delay prediction. The focus is on design decisions, business reasoning, and model choices rather than mathematical complexity.

1. Attribution Modeling Logic

The marketing and sales datasets in this project operate at different levels of granularity. Marketing data is recorded at a daily level, while orders data is transactional with multiple records per day. Because there is no direct campaign-level mapping between orders and marketing campaigns, traditional attribution methods such as last-click attribution are not feasible.

To address this challenge, a time-series attribution approach was adopted. Orders data was first aggregated by date to compute total daily revenue. This aggregated revenue was then joined with daily marketing spend using the date as the common key. This approach resolves the granularity mismatch and allows a fair comparison between marketing investment and generated revenue.

The primary metric derived from this attribution model is ROAS (Return on Ad Spend), calculated as total revenue divided by total marketing spend. ROAS provides a clear indicator of marketing efficiency and helps identify profitable and non-performing marketing periods.

2. Machine Learning Model and Hyperparameters

The machine learning component focuses on predicting delivery delays before they occur. This is framed as a binary classification problem, where the model predicts whether an order will be delivered late or on time.

Feature engineering was performed by extracting time-based features from the order timestamp, including hour of day and day of week. These features capture traffic patterns and peak delivery periods. Geographic features such as region were not available in the dataset and were therefore excluded.

A Logistic Regression model from scikit-learn was chosen as a baseline classifier due to its simplicity, interpretability, and ability to output probability scores. These probabilities are used in the application to classify deliveries into low, medium, or high risk categories.

The model was trained using default regularization settings with an increased maximum number of iterations to ensure convergence. Model performance was evaluated using the Area Under the ROC Curve (AUC), which measures the model’s ability to distinguish between late and on-time deliveries. A lower AUC score reflects limitations in available features rather than deficiencies in model selection.

3. Conclusion

This project demonstrates how analytics, machine learning, and generative AI can be combined to support data-driven decision-making in a quick-commerce environment. The attribution model provides transparency into marketing effectiveness, while the predictive model enables proactive operational planning based on delivery risk.