

XGBoost Model for Fraud Detection

Crank That Bayesian Handle

July 10, 2025



Why We Chose XGBoost

We chose XGBoost for its strong performance on structured, tabular data like transaction records. It handles non-linear relationships and feature interactions effectively, often outperforming simpler models like logistic regression without requiring heavy parameter tuning. XGBoost also includes built-in support for handling imbalanced datasets, which is often the case in fraud detection problems.

Feature Engineering: Additional Features

To enhance the model's predictive power, we engineered additional features capturing temporal and behavioral patterns of transactions:

- **Time Since Last Transaction:** For each transaction, we calculated the time elapsed since the previous transaction on the same card. This helps capture unusual transaction timing patterns that might indicate fraud.
- **Amount Difference from Last Transaction:** We computed the difference in amount between the current and the previous transaction on the same card, providing insight into abnormal spending behavior.
- **Z-score Anomaly for Amount:** We computed the Z-score of each transaction amount relative to the historical mean and standard deviation of amounts for the same card. This standardizes the amount and flags statistically significant outliers.
- **5-Day Rolling Mean of Amount:** We calculated the rolling average transaction amount over a 5-day window per card. This smooths out short-term fluctuations and captures changes in spending trends.
- **Cumulative Number of Transactions per Card:** For each transaction, we tracked the cumulative count of previous transactions made with the same card. This captures transaction volume and usage patterns over time.