SGH x Mastercard Hackathon - May 2025

Report

Team: ***LAMP***

Laura Hoang, Antoni Ballaun, Mateusz Kalinowski, Piotr Bielecki

1. Project Objective

The primary objective of this project was to build a classification model capable of detecting fraudulent transactions. Due to the significant financial impact of fraud, even partial detection can improve business security and reduce losses. We framed this as a binary classification task, where:

- Class 0 represents non-fraudulent transactions
- Class 1 represents fraudulent transactions

Given the sequential nature of transactional data, we chose a Long Short-Term Memory (LSTM) neural network, which is effective at capturing temporal dependencies and user behavior over time.

2. Dataset and Features

The dataset contains 500,000 samples, each reflecting transactional behavior. Features included:

- Numerical time series (e.g., amounts, user age)
- Time-based features (e.g., timestamps, signup dates)
- Categorical features (e.g., payment method, currency, channel)
- Binary flags (e.g., international transaction, previous fraud history)

The dataset was highly imbalanced, with approximately 8.5% of transactions labeled as fraudulent.

3. Data Preparation and Modelling Approach

The data was split into training and validation sets using stratified sampling to preserve class distribution. Categorical features were encoded using embedding layers within the model and numerical features were normalized to improve training stability. New features where calculated based on given ones.

The model architecture combined:

- Three stacked LSTM layers with Layer Normalization for sequence modeling
- Embedding layers for categorical features
- Dropout layers for regularization
- A final Dense layer with sigmoid activation for binary classification

Before feeding the data into the LSTM, individual transactional observations were grouped into fixed-length sequences per user, enabling the model to learn behavioral patterns over time.

The model contained approximately 1.25 million parameters, with around 415,000 trainable.

4. Results

The trained model achieved the following performance on the validation set:

- Accuracy: 73.6%
- Precision (fraud class): 11%
- Recall (fraud class): 30%
- F1-score (fraud class): 0.16
- Weighted F1-score: 0.79

Confusion Matrix:

	Predicted Non-Fraud (0)	**Predicted Fraud (1)**
Actual Non-Fraud (0)	70,987	20,531
Actual Fraud (1)	5,911	2,571

Despite the dataset's imbalance, the model was able to correctly identify 30% of fraudulent cases, demonstrating a promising baseline for further improvement.

5. Challenges and Observations

Several challenges affected model performance:

- Class imbalance: Fraud cases were rare, leading to a bias toward the majority class.
- Limited fraud signals: The available features lacked richer contextual information (e.g., user IP, device fingerprints), which likely reduced detection power.
- Feature engineering difficulties: Crafting effective features to highlight subtle fraud patterns was challenging, limiting the model's ability to generalize well.

6. Conclusion

This project demonstrates that an LSTM-based model can capture temporal patterns in transactional data and provide a foundation for fraud detection despite inherent difficulties like class imbalance and feature limitations. With further enhancements — including advanced sampling strategies, threshold optimization, and enriched feature engineering — this approach can be developed into a robust fraud detection system. Despite challenges with feature engineering, some still proved to be effective – mostly the ones related to time and location turned out to be the most predictive.