Credit Card Fraud Detection Using Deep Neural Networks

Elizabeth Obst
Olivia Shipley
Shamir Cardenas
Jordan Bona

IS 6733 - Fall 2025

Introduction and Problem

Credit card fraud poses a significant threat to the financial industry, resulting in billions of dollars in annual losses globally. Traditional rule-based fraud detection systems often fail to adapt to evolving fraud tactics and generate excessive false positives that burden investigation teams and degrade customer experience. These systems struggle to adapt to increasingly sophisticated fraud schemes.

This research evaluates the effectiveness of deep neural networks in detecting fraudulent credit card transactions using the Kaggle Credit Card Fraud Detection dataset, which contains 284,807 European credit card transactions from September 2013. The dataset presents the typical challenge faced in real-world fraud detection: extreme class imbalance with only 492 fraudulent transactions (0.172%) among 284,315 legitimate transactions (99.828%). The primary objective is to develop a model that accurately identifies fraudulent transactions with high precision while maintaining high recall to catch the majority of actual fraud cases.

Challenges

This fraud detection problem presented several challenges with the most prominent being extreme class imbalance. With only 492 fraudulent transactions (0.172%) among 284,315 legitimate transactions, this can cause models to predict all transactions as legitimate to achieve high accuracy. Due to our dataset being credit card transactions, all of our features were anonymized, 28 principal components (V1-V28) were created through PCA transformation to protect customer privacy, this precaution limited interpretability but maintained the data utility.

A traditional metric such as accuracy can be misleading in some circumstances, this made choosing appropriate evaluation metrics difficult. Accuracy can be problematic because a model predicting all transactions as legitimate would achieve 99.8% accuracy while detecting zero fraud. To combat this, we chose to use precision and recall as the appropriate metrics. This solution was not without issue, we had a hard time balancing the two metrics. Fraud detection requires catching fraudulent transactions (recall) while minimizing false positives, as a result, this can waste investigation resources and frustrate legitimate customers (precision). We will address our solutions below.

Proposed Solution

We developed a deep neural network architecture with four hidden layers designed to learn hierarchical representations of fraud patterns. The network consists of an input layer accepting 30 features (V1-V28 principal components plus Time and Amount), followed by hidden layers with 64, 32, and 16 neurons respectively, each using ReLU activation. Dropout regularization (30%, 30%, 20%) was incorporated at each hidden layer to prevent overfitting, and the output layer uses sigmoid activation for binary classification.

To address the extreme class imbalance, we applied Synthetic Minority Over-sampling Technique (SMOTE) exclusively to the training data, generating synthetic fraudulent transactions by interpolating between existing minority class instances. This approach increased fraud representation to approximately 50% of the training set while preserving the original imbalanced test set for realistic evaluation. The model was trained using the Adam optimizer with binary cross-entropy loss for 20 epochs with a batch size of

256. Without SMOTE, the neural network would simply learn to classify all transactions as legitimate to achieve high accuracy, rendering the model ineffective for fraud detection.

Methodology

The experimental approach consisted of data preprocessing, addressing class imbalance through SMOTE, neural network architecture design, and systematic comparison against baseline models. The dataset was split into 80% training (227,845 transactions) and 20% test sets (56,746 transactions) using stratified sampling to maintain the original fraud ratio. StandardScaler normalization was applied to Time and Amount features to ensure consistent scaling across all inputs to the neural network.

Three model families were implemented for comprehensive comparison: a deep neural network with dropout regularization, XGBoost with class weight balancing, and Logistic Regression with balanced class weights as an interpretable baseline. All models were evaluated using precision, recall, F1-score, and Precision-Recall Area Under Curve (PR-AUC) as the primary metrics. Traditional accuracy was avoided as it would be misleading for imbalanced data.

Evaluation and Results

Model Performance Summary

The deep neural network achieved strong performance with 83% precision and 72% recall (F1-score: 0.768), successfully identifying 68 of 95 fraudulent transactions in the test set while generating only 14 false alarms. This represents competitive performance, particularly in precision where the DNN slightly outperformed XGBoost. The PR-AUC of approximately 0.744 indicates strong performance across different classification thresholds.

XGBoost emerged as the best-performing model with the highest F1-score (0.802) and PR-AUC (0.811), demonstrating superior balance between precision and recall. The model achieved 82% precision and 79% recall, successfully identifying 75 of 95 fraudulent transactions while generating 17 false alarms. The key advantage of XGBoost was its better balance between catching fraud and minimizing false alerts. These results align with established machine learning literature showing that gradient boosting methods often outperform neural networks on structured tabular data, particularly with datasets of this size (280K transactions).

Logistic Regression, while achieving the highest recall (87%), produced an impractical number of false alarms (1,386 false positives), resulting in only 6% precision. This translates to approximately 25 fraud alerts per 1,000 transactions, of which only 1-2 would be genuine fraud. Such performance would overwhelm fraud investigation teams and is unsuitable for production deployment.

Business Context

Fraud detection requires balancing two objectives: catching fraudulent transactions (recall) while minimizing false positives (precision). False alarms waste investigation resources and may frustrate legitimate customers, while missed frauds result in direct financial losses. For context, the neural network generated only 14 false positives but missed 27 fraudulent transactions, while XGBoost produced 17 false alarms but missed only 20 frauds, illustrating the trade-off between precision and recall.

While XGBoost outperformed the neural network on this dataset, deep learning excels in production scenarios with larger-scale data (millions/billions of transactions), temporal patterns (RNNs/LSTMs for transaction sequences), multi-modal data (text, images, geographic data), and complex non-linear feature interactions. The Kaggle dataset's 280K pre-processed transactions with anonymized features represents a simplified scenario; production systems with rich, multi-dimensional data favor neural networks' representation learning capabilities.

Conclusion

This research demonstrates that deep neural networks are highly effective for credit card fraud detection. The DNN achieved competitive results with 83% precision surpassing XGBoost's 82%, which is a critical advantage for production systems where investigation resources are limited and false alarms are costly. While XGBoost achieved slightly better overall performance (F1-score: 0.802, PR-AUC: 0.811) due to higher recall, the DNN's results validate deep learning's promise for real-world deployments, particularly with larger datasets, temporal sequences, and multi-modal data sources.

Future enhancements could include: (1) RNN/LSTM architectures to capture temporal transaction patterns, (2) ensemble approaches combining multiple network architectures, (3) attention mechanisms to focus on relevant features, and (4) training on larger datasets with additional features like merchant categories and geographic data. In conclusion, while tree-based methods remain pragmatic for many scenarios, neural networks represent a powerful fraud detection tool that will become increasingly valuable as data volume and complexity grow.