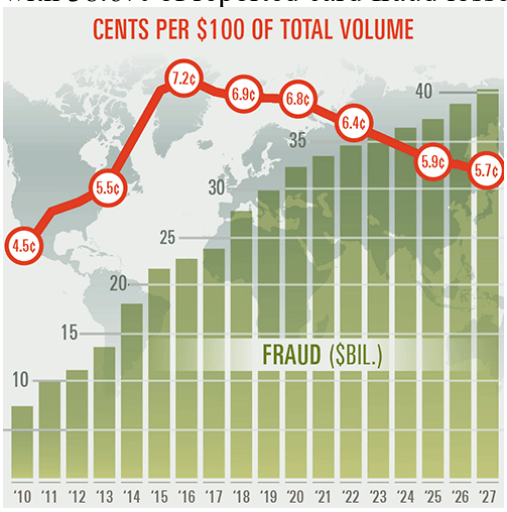
Credit Card Fraud Detection Using Machine Learning

Any time someone uses any sort of cards such as debit, credit to make purchases as opposed to using cash or checks, falls under card-based payment category. The world has witnessed a tremendous growth of card-based transaction volume in recent years. By 2022, card-based payment will occupy 83% of total transactions in USA.

Credit Card fraud i.e. unauthorized access of someone's credit card, is one of biggest problem that financial industries are facing since introduction of card-based payments. Credit Card fraud is one of the biggest sources of revenue loss for payment card industry. By federal law, individuals are protected from transaction fraud, but the loss has to incurred by either financial institution or merchant payment processor. In 2018, $24.26 Billion was lost due to payment card fraud worldwide. The United States leads as the most credit fraud prone country with 38.6% of reported card fraud losses in 2018.



Financial institutions use traditional fraud detection system. **Traditional Fraud Detection** system is rule based where fraud analysts writes rules to catch fraud. Highly tenured fraud analysts analyze the historical transaction details and confirmed fraud cases to determine the fraud trend and then writes rules to catch these frauds. Then these rules are deployed into production to combat fraud. Fraudsters find new ways to bypass these rules and commit transaction fraud. Analysts re-analyze the new fraudulent trend to update the rules. There is a constant battle between fraud analysts and fraudsters.

Some issues with traditional fraud detection system:

- Requires domain expertise to build and maintain the system.
- High operating cost for maintenance.
- Greater Time to market to combat new fraud trends.

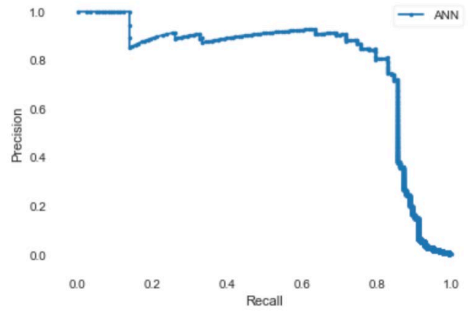
Due to these issues, financial institutions are looking for some alternatives. Machine learning can come into play to provide suitable alternatives. There are various machine learning classification algorithms such as Random Forest, Artificial Neural Network, Logistic Regression etc. that can be used to detect credit card fraud. These models can be trained offline using historical transaction details and confirmed fraud data. These model's performance can be

compared based on their capabilities to detect genuine fraudulent transactions v/s improperly classifying legitimate transactions as fraudulent. Once the best model is identified and trained, then it can be deployed into production to combat fraud.

The current project prototype used two days historical transactions from 2013 happened in Europe. It contains 284,807 transactions out of which 492 are fraudulent. Exploratory data analysis was performed first to understand the relationship between different transaction attributes and then the data was cleaned. Using a cross validation technique, the different model parameters were tuned and then compared.

Model	Best Average Percision Score
Random Forest	0.8785
Logistic Regression	0.8716
Artificial Neural Network	0.8770

Best average precision score signifies efficiency to genuine fraud transactions as compared to all transactions classified as fraudulent. Higher the number better is the efficiency. Artificial Neural Network model was selected as a best model after considering calibration and complexity. Finally, it was tested against an unseen sample from the dataset and it performed well with average precision-recall score 0.79.



Higher area under this precision-recall curve signifies that this model will do well to detect fraud from generalized credit card transactions. This prototype of Artificial Neural Network model can be deployed into production to detect fraud and tuned further based on the confirmed fraud data from production.

The advantages of Machine Learning Fraud Detection System over Traditional Fraud Detection System:

- Less human involvement.
- No need for domain expertise.
- Once the model is built and deployed into production, very little operating cost.
- Periodical re-training can combat newer fraud trends.

Quick time to market, efficient, less human dependency, low operating cost and greater adaptability, makes Machine Learning an easy and good choice to replace the traditional fraud detection systems.