Credit Card Fraud Detection

Credit card fraud is a term that has been coined for unauthorized access of payment cards like credit cards or debit cards to pay for using services or goods. Hackers or fraudsters may obtain the confidential details of the card from unsecured websites. When a fraudster compromises an individual's credit/debit card, everyone involved in the process suffers, right from the individual whose confidential data has been leaked to the businesses (generally banks) who issue the credit card and the merchant who is finalizing the transaction with purchase. This makes it extremely essential to identify the fraudulent transactions at the onset. Financial institutions and businesses like e-commerce are taking firm steps to flag the fraudsters entering the system. Various advanced machine learning technologies are at play, assessing every transaction and stemming the fraud users in its nip using behavioral data and transaction patterns. The process of automatically differentiating between fraudulent and genuine users is known as **"*credit card fraud detection*".**

A credit card is one of the most used financial products to make online purchases and payments such as gas, groceries, TVs, traveling, shopping bills, and so on because of the non-availability of funds at that instance. Credit cards are of most value that provide various benefits in the form of points while using them for different transactions. There are several categories of credit card fraud that are prevalent in today's time:

- **Lost/Stolen cards:** People steal credit cards from the mail and use them illegally on behalf of the owner. The process of blocking credit cards that have been stolen and re-issuing them is a hassle for both customers and credit card companies. Some financial institutions keep the credit cards blocked until it is verified that the rightful owner has received the card.
- **Card Abuse:** The customer buys goods and items on the credit card but has no intention to pay back the amount charged by the bank for the

same. These customers stop answering the calls as the deadline to settle the dues approaches. Sometimes they even declare bankruptcy—this type of fraud results in losses of millions every year.

- **Identity Theft:** The customers apply illegitimate information, and they might even steal the details of a genuine customer to apply for a credit card and then misuse it. In such cases, even card blocking can not stop the credit card from falling into the wrong hands.
- **Merchant Abuse**: Some merchants show illegal transactions (that never occurred) for money laundering. For performing these illicit transactions, legal information of genuine credit card users is stolen to generate replicas of the cards and use it for illegal work.

Credit Card Fraud Detection using Machine Learning can be done using:

- **Unsupervised Learning** - Machine Learning Algorithms such as Isolation Forest, One-class SVM, LOF, etc., do not require labeled data for training the model. They identify patterns in the data and try to group the data points based on observed similarities in patterns.
- **Supervised Learning** - Machine Learning Algorithms such as Ensemble Models (RandomForest, XGBoost, LightGBM, etc.), KNN, Neural Networks, Autoencoders, etc. These algorithms are trained on labeled data, and the model learns to predict the labels for the unseen data.

Challenges in Credit Card Fraud Detection:

The challenges involved in credit card fraud detection project is primarily the data itself. The data is heavily imbalanced, i.e., the count of data labeled as fraudulent is way less than the data labeled as non-fraudulent data. This makes it extremely tricky to train the model as it tends to overfit for the majority class and underfit for the minority class. Techniques like oversampling, undersampling, cost-sensitive learning, etc. can be used to deal with this. The metrics used for the final model are different from standard evaluation metrics of accuracy, AUC-ROC, etc.

Another prevalent faced challenge is the quality and quantity of the data. The startups in the early stage do not have much user history data to train extensive models, which makes it difficult to train a robust fraud detection model. A temporary solution to this problem can be sourcing data from an external third party, like scores from credit bureaus.

Conclusions:

We learned how to develop our credit card fraud detection model using machine learning. We used a variety of ML algorithms, including KNNs and Tree-based models. At the end of the training, out of 8544 validation transaction.