Credit Card Fraud Detection Using Machine Learning

Nausheen Samiha

Department of Computer Science and Engineering,

School of Data & Science,Brac University.

Introduction: Credit card is one of the most accessible method for money transaction and payments in the modern days. Every single financial institution and banks have credit card for the citizens and foreign business persons. With the increasing popularity of credit card usage the number of frauds have also increased in every single bank like not paying off the card and having multiple cards for the scamming purpose. However, it is very hard to detect the frauds with the conventional method of banking. In this paper, we have used machine learning based models to predict the future possible frauds depending on their previous activity. Machine Learning models like CNN and LSTM is trained and tested on public data to detect the nature of the frauds, early signs and the changing data trends.

Background: Credit card fraud cases has increased in Bangladesh in recent years tremendously. In a reputed bank, there has been a scam of 10 crore through 21 credit cards and recently, there has a fraud case of 3 crores showing false transaction in the ATM booth .Though the customers are not facing any problem, the financial institutions are facing serious losses. There is currently $120 billion worth of credit card holders in Bangladesh and to measure such value machine learning algorithms are used . Machine Learning, A specific part of artificial intelligence has used the models like auto encoders, Long Short Term Memories, Convolution Neural Network.Also ANN and Decision trees are used for fraud detection with small dataset and less complex data. A hybrid model of those is also used for predictive analytic method.

Plans: In the past, ANN,Decision trees and auto encoders have already been used and they are suitable for small datasets. For large datasets, CNN and LSTM is used to train huge datas within the least amount of time. As data preprocessing is an important stage in ML, CNN and LSTM can complete it accurately.Just like ANN, convolutional neural

network has 3 layers :input layer, hidden layer and output layer with different number of channels in the each layer.CNN are not fully connected, each layer has fixed weight and different number of channels, kernel size, activation function, flattening which makes it faster in calculation.LSTM is a part of RNN which minimises the loss while adjusting the weight.In this paper, the goal is to detect the performance of classifier from dataset with various number of samples and features.with the combination of LSTM after CNN, all of the datasets will take the time, amount of transaction, foreign transaction, declined amount, average amount per day, merchant ID and place. After getting all the data, it will use accuracy, precision, recall to measure the data preprocessing, data scaling and standardisation. In the training set 80% data is taken and rest 20% is in the testing phase.If the result is 0 then it's a valid customer and if the result is 1 it's a possible fraud.

Potential Challenges: to detect fraud ,we need to get all the values of the customers.However, sometimes the informations are confidential and can not be shared .Without proper data, it is possible to get a wrong outcome and the model can not get all the possible frauds without enough data and that's why many times the predicted outcome is way different then the real world fraud case numbers. Also with this model there is a high chance of getting false negative or false positive that can mislead the result. Moreover, the data can be messy ,overwritten with a lot of missing information which leads to data cleaning for a longer time.

Conclusion: Everywhere, frauds are coming with different methods to create fake transaction. So, there is also changes in data every-time we update the dataset.In ML ,we do not have to change the algorithm time to time, it will adjust itself.This model can predict almost 85% accurate result with all the related info. As CNN is used for image processing, it can produce the result quickly than ANN and decision trees.To get more precise result, the ML algorithms can be combined into hybrid one and in the future there can be more hybrid architecture to create efficient outcome.