Credit Card Fraud Detection Using Machine Learning

Nausheen Samiha
Department of Computer Science and Engineering,

School of Data & Science,

Brac University.

Abstract: This paper shows the theory for creating an automated Credit Card Fraud detecting method to implement in the real world money transaction. Almost all Citizens especially young generations utilises credit card more than cash money for the safety purpose. However, it is a risk for the financial companies as it often fails to detect the fake transaction happening nearly in every financial companies. Previously ML algorithm ANN, bayesian network , auto encoder has been used which has a lot of faulty results.So, In this paper CNN and LSTM is proposed together as a hybrid network to detect the possible fraud customers and transactions more accurately.

1.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 Convolution Neural Network(CNN) and Long Short Term Memory (LSTM) is trained and tested on public data to detect the nature of the frauds, early signs and the changing data trends.After the Training of the ML algorithm it will be able to detect the frauds in advance.

2.Background

In this section, we will discuss why credit card fraud detection is an important task for any financial institution of Bangladesh. Number of credit card fraud cases has increased in Bangladesh in recent years tremendously. In a reputed bank, there has been a scam of 10 crores taka through 21 credit cards by a group of foreign scammers and recently in 2022, there has been a fraud case of 3 crores showing false transaction in the ATM booth .The case of this year has been committed by a group of IT professional of an security firm.That's why, it's really important for every large institution to have previously built ML models by their own employees to detect upcoming fake customers.Again, Stealing the credit card security number from many financial transactions and using it to withdraw a fortune amount is a very common is in Bangladesh and for this case, the bank or the financial company has to cover up for their customers. 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 is a specific part of artificial intelligence that is able to learn without being programmed in details and it almost acts like a human brain which takes previous data to learn and autocorrects the mistakes to create the best fitted output. Machine Learning has used the models like auto encoders, Long Short Term Memories, Convolution Neural Network.Also Artificial Neural Network 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. ANN,CNN And RNN (LSTM) is a part of supervised learning where the data is labeled and this kind of learning is necessary for understanding the customer data accurately.

3.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 : an input layer which takes data from the database or from a device, a hidden layer which does not have any interaction with outside of

the model and an output layer which shows two or more output as needed with different number of channels in the each layer.CNNs 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 Recurrent Neural Network(RNN) which minimises the loss while adjusting the weight in each epoch. Epoch is the number that each algorithm goes through an entire training phase to adjust the weight from each learning. LSTM remembers specific value or forgets it depending on the importance and so it the customer is not a threat the algorithm will specifically choose a value to keep the minimum use of memory. 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 getting the input from CNN's output layer and 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.After calculating accuracy, precision and recall If there is not any closer value generated than the result is 0 indicating that it's a valid customer and if the result is 1 it's a possible fraud.

4.Potential Challenges

 The number of credit are users are increasing for it's comfort usage and safety benefits from the banks. To detect a genuine 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 than 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. For the ML model it must be able to handle noises while generating a very precise data because the number of credit card fraud is way less than other fraud cases.If it fails to give that

precision, many cards can be blocked from false positive result. There has to be different training and testing for different currencies like dollar and euros as credit cards are the only source for international purchase and it will be very time consuming to create different training and combining them for a single customer. Even though CNN and LSTM is a very fast learning system, while handling big data it will take 2 to 3 days to complete the entire training phase and thus it is not the quickest algorithm out there.

5.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 with the new inputs it takes every time.This model can predict almost 85% accurate result with all the related info. As CNN is used for image processing and LSTM is often used in creating prediction for next few weeks, 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.

Reference:

1.Credit card fraud detection using Bayesian and neural networks : https://www.researchgate.net/profile/Karl-Tuyls/publication/254198382_Machine_Learning_Techniques_for_Fraud_Detection/links/555f695508ae6f4dcc926e88/Machine-Learning-Techniques-for-Fraud-Detection.pdf

2.Deep Learning Methods for Credit Card Fraud Detection-Thanh Thi Nguyen,Hammad Tahir,Mohamed Abdelrazek,Ali Babar -https://arxiv.org/abs/2012.03754

3.ATM forgery unearthed-https://www.thedailystar.net/news/atm-forgery-unearthed

4.Detecting Credit Card Fraud using Machine Learning-https://online-journals.org/index.php/i-jim/article/view/27355