

**Credit Card Fraud Detection Using Historical Transactions**

**(Project term jan-may 2020)**

**Artificial Intelligence**

Submitted by

Sai Sathya **11815101**

**Course code: INT 404**

Submitted to

Mr.Sagar Pande

UID : 23754

**School of computer science and engineering**

TABLE OF CONTENT

* Abstract
* Introduction
* Implementation
* Conclusion
* Reference

ABSTRACT:

The project which we have took up is credit card fraud detection using historical data. In this project we are supposed to detect fraud transaction which happened in credit card.

* With the growth of e-commerce websites, people and financial companies rely on online services to carry out their transactions that have led to an exponential increase in the credit card frauds [1]. Fraudulent credit card transactions lead to a loss of huge amount of money. The design of an effective fraud detection system is necessary in order to reduce the losses incurred by the customers and financial companies.
* Research has been done on many models and methods to prevent and detect credit card frauds. Some credit card fraud transaction datasets contain the problem of imbalance in datasets. A good fraud detection system should be able to identify the fraud transaction accurately and should make the detection possible in real-time transactions. Fraud detection can be divided into two groups: anomaly detection and misuse detection. Anomaly detection systems bring normal transaction to be trained and use techniques to determine novel frauds. Conversely, a misuse fraud detection system uses the labeled transaction as normal or fraud transaction to be trained in the database history. So, this misuse detection system entails a system of supervised learning and anomaly detection system a system of unsupervised learning.
* Fraudsters masquerade the normal behavior of customers and the fraud patterns are changing rapidly so the fraud detection system needs to constantly learn and update. Credit card frauds can be broadly classified into three categories, that is, traditional card related frauds (application, stolen, account takeover, fake and counterfeit), merchant related frauds (merchant collusion and triangulation) and Internet frauds (site cloning, credit card generators and false merchant sites).

# Introduction

Motivation :

Fraud detection involves in the data manipulation, applying algorithms and

help to gain knowledge about credit card fraud detection using historical transaction

data &cyber security.

It is good project to get knowledge about credit card fraud detection.

Aim :

The aim of this project is to detecting the credit card frauds transaction using historical data.

Objectives:

Cashless transactions such as online transactions, credit card transactions, and mobile wallet are becoming more and

more popular in financial transactions nowadays. With increased number of such cashless transaction, fraudulent transactions are also

increasing. Fraud can be detected by analyzing spending behavior of customers (users) from previous transaction data. If any deviation

is noticed in spending behavior from available patterns, there may be chance of fraudulent transaction. To detect fraud behavior, bank

and credit card companies are using various methods of data mining such as decision tree, rule based mining, neural network, fuzzy

clustering approach, hidden markov model or hybrid approach of these methods. In this paper we have used Convolutional neural

network with SMOTE. We have transformed original features into new features.

What is credit card fraud detection:

The credit card fraud detection problem includes modelling past credit card transaction with the knowledge of the once that turned out to be fraud. This is a model is then used to identity whether a new transaction is a fraudulent or not.

Problem with credit card fraud detection:

One of the biggestproblem associated with researchers in fraud detection is lack of

real life data because of sensitivity of data and privacy issue. Many researchers have

done research with real life data [3], [9], [7], [11] of bank with agreements. To deal

with this problem, many tools are available to generate synthetic data.

Second problem is to deal with Imbalance data or skewed distribution because number

of fraudulent transactions are very less compare to legitimate transactions. To overcome this problem, synthetic minoring oversampling methods are used to increase number of low incidence

data in dataset that generate synthetic fraudulent transactions related with original data set.

In [3], cost

based sampling is used to generate synthetic fraudulent transactions to balance data set.

Overlapping of data is another problem as some of transactions

look like fraudulent transaction,when actually they are

legitimate transactions. It is also possible that fraudulent transactions appear to be normal transactions.

Implementation :

There are several effective methods to detect banking transaction frauds. Depending on innovative transaction procedures used by

frauds, these methods may fail in detecting fraudulent transaction and may cause enormous damage to Card issuers or users. Here,

by adding new features in dataset and SMOTE sampling method improves results using Convolutional Neural Network by

detecting outlier transactions which can be fraudulent transaction of credit card usage.

Our proposed flow work is divided into two parts.

1. Training phase

2. Prediction phase

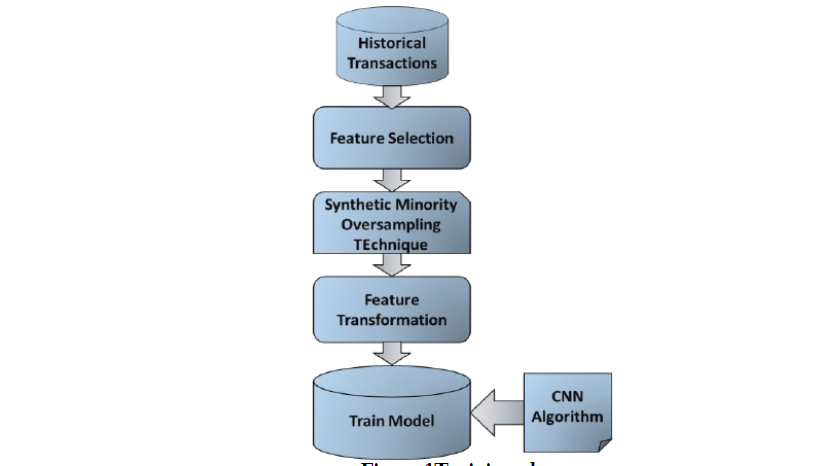
In training phase, we will give historical transactions as an input included legitimate and fraudulent. In training phase, Feature

Selection of attribute is done and then Synthetic Minority Oversampling Technique (SMOTE) method will be applied to generate

synthetic frauds to overcome issue of imbalanced data. We have introduced some features that can be generated from raw

features. In order to apply CNN model, we need to transform features into feature matrix to fit the model.

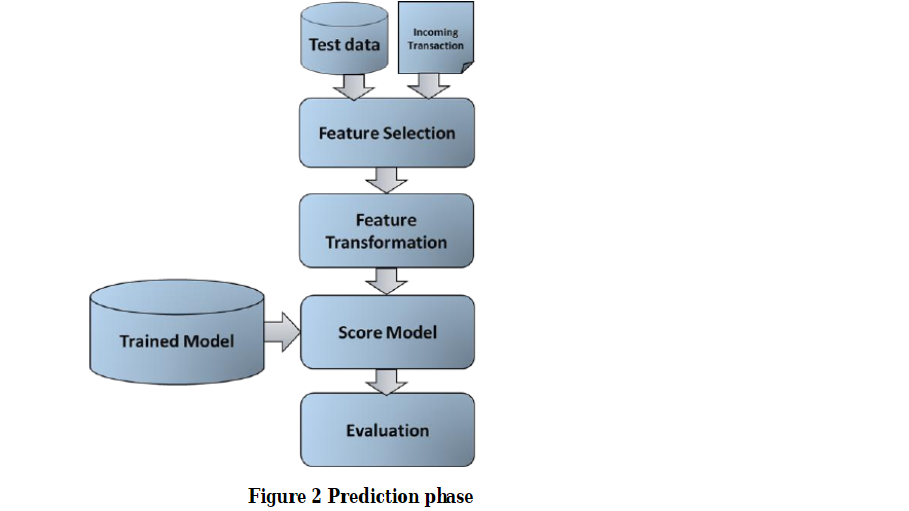
Then we will train CNN model.



In Prediction phase, when new transaction will come, it will be given as input. After new transaction’s feature extraction and

transformation, it will be tested with our CNN trained model classifier. It will be resulted as fraudulent transaction or legislative

transaction.



A.Feature selection

Feature extraction is used to extract useful features to train the model from dataset. These features are extracted from raw data.

User id

Date of transaction

Merchant

Amount

Fraudulent or not

B.SMOTE

One of the problem in fraud detection is to deal with Imbalance data or skewed distribution because number of fraudulent

transactions are very less compare to legitimate transactions. To overcome this problem, synthetic minoring oversampling

technique is used to increase number of low incidence data in dataset that generate synthetic fraudulent transactions related with

original data set.

In [13], SMOTE is used to generate synthetic fraudulent transactions to balance data set. Using this technique, fraudulent samples

will be increased, those samples will be given with original transactions to train model.

C.Feature transformation

After selecting features from dataset, pre-processing on data is done and these features are converted as average amount, total

amount, number of transactions, bias and trading entropy with reference to current transaction on previous data[3]. For this time

window of three day, one week, fifteen days, one month and from beginning of account is taken as shown in figure 3.

Here we have introduced a new feature named bias with merchant.

Avg\_Amount\_T : Average amount of transactions during past period of time.

NumberT : Total number of the transactions during the past period of time.

TotalAmountT : Total amount of the transactions during the past period of time

BiasAmountT : The bias of the amount of this transaction and AvgAmountT

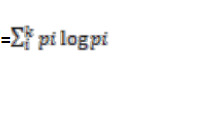
Trading Entropy: Assume in all transactions of the same customer during the past period of time before the current transaction,

there are K kinds of merchant types, the total amount is TotalAmountT, the sum amount of the i-th merchant type is AmountTi(i

= 1, 2, . . .,K), the proportion of the i-th merchant type is pi:

pi = amounti/totalamount (1)

The entropy of the i-th merchant type can be defined as EntT:

EntT=  (2)

The above calculations only use previous transactions while the current transaction is not involved in. Then we add the current

transaction to join the above calculation to obtain the current entropy: NewEntT. So the trading entropy is defined as

TradingEntropyT:

TradingEntropyT = EntT-NewEntT (3)

Merchant Bias: Assume in all transactions of the same customer during the past period of time before the current transaction.

Suppose current transaction is done with merchant x with amountX. Merchant X’s average amount during previous transactions’ is AvgamountX.

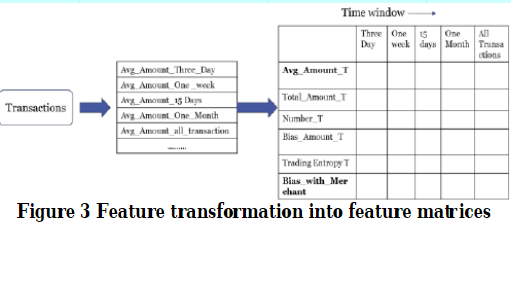
So bias amount of amountX and AvgamountX is defined as bias\_with\_merchant. So for i-th merchant in current transaction, bias\_with\_merchent = amounti–Avgamounti (4)

First these features are converted into one dimension, then converted into matrix while training CNN model.

If history of customer is not available, i.e. for a new customer, all values for Avg\_amount\_T , Total\_amount\_T, Number\_T, Entropy\_T

this issue, we averaged all customers who have done transaction in

time period T and thenhave taken bias with thataverage amount.

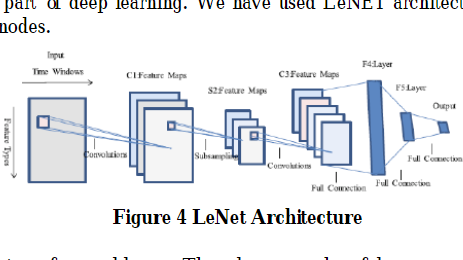


D.Train Model

For training a model we have used convolutional neural network (CNN).

This converted features will be given as input to CNN.

Convolutional Neural Network is a part of deep learning. We have used LeNET architecture of CNN. We have used softmax function for neuron and 100 hidden nodes.



A convolutional neural network consists of several layers. These layers can be of three types.

Convolutional: Convolution means mapping of input layer’s neuron to hidden layer (Convolution layer). Each neuron of

convolutional layer takes input from previous layer and convolute them to convolutional layer. Taken input should be rectangular

grid of neurons. In this, weights specify filter of convolution.

Sub-sampling: Sub sampling layer makes different samples of convolutional layer that is also called as feature map. Subsampling

also reduces parameters from previous layer. Average, Maximum are normally used functions for sub sampling.

Fully-Connected: Fully connected layer takes all neurons of previous layer and connect it to every single neuron. After fully

connected layer, no convolution is possible.

V.MATRICS TO EVALUATE SYSTEM

As the data is highly imbalance, overall accuracy is not appropriate to evaluate model, since with very high accuracy, almost all

fraudulent transactions can be misclassified.

Precision, recall, F1 score, Ratio of True Positive, True Negative, False Positive and False Negative are taken into account for

evaluating binary classification.

True Positive (TP) is number of correctly classified fraudulent transactions.

True Negative (TN) is number of correctly classified legitimate transactions.

False Positive (FP) is number of incorrectly classified legitimate transactions.

False Negative (FN) is number of incorrectly classified fraudulent transactions.

Precision (P) = TP/TP+FN (5)

Recall (R) = TP/TP+FN (6)

F1 score is harmonic mean of precision and recall. Value of F1 score lies between 0 to 1. Higher F1 score indicates good model.

F1 score = 2\*PRECISION+RECALL/PRECISION+RECALL (7)

VI.EXPERIMENT

A.Description of dataset

In my research data set used is synthetic data set as real credit card transactions data set is not available due to privacy of

customers. I have generated data from [12].

Recourse: https://github.com/metasyn/creditcardfrauddata

Instances: 21300

Attributes:

1.Date

2.Time

3.User id

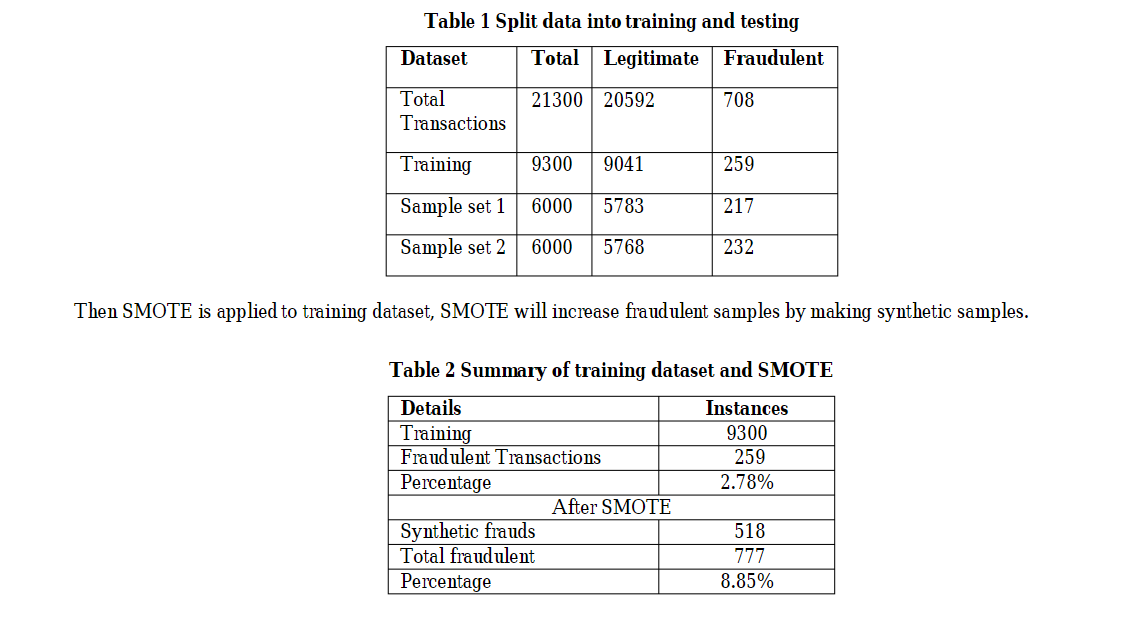
4.User name

5.Merchant

6.Amount

7.Fraudulent

First this data is split into training set and testing.



B.Environment

I have used Microsoft azure machine learning studio for implementing CNN classifier. To define CNN, used Net# provided by

Azure ML. Azure Machine Learning Studio is a GUI-based integrated development environment for constructing and operationalizing Machine Learning workflow on Azure. I have also used R studio for feature transformation. This transformed

features are given as input to CNN classifier to train model.

C.Results

We have experimented using NN and CNN both and then compared results of both.

P= precision

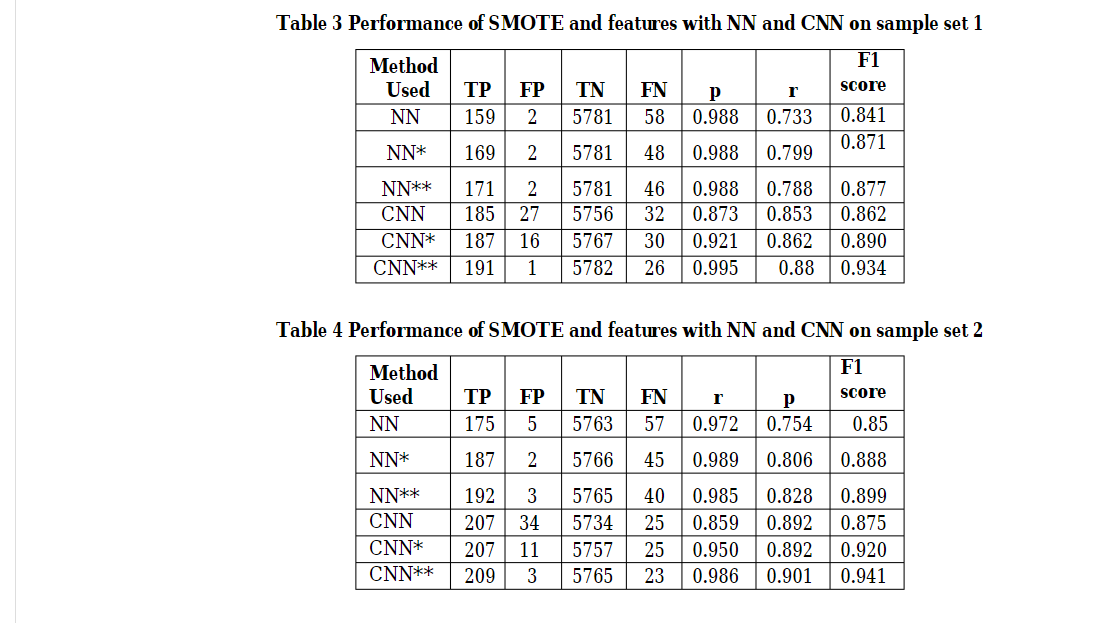
R= recall

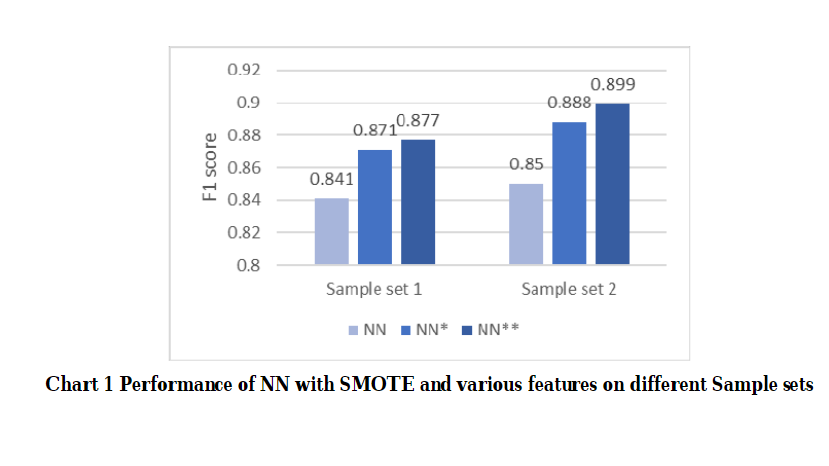
NN\* = NN with SMOTE

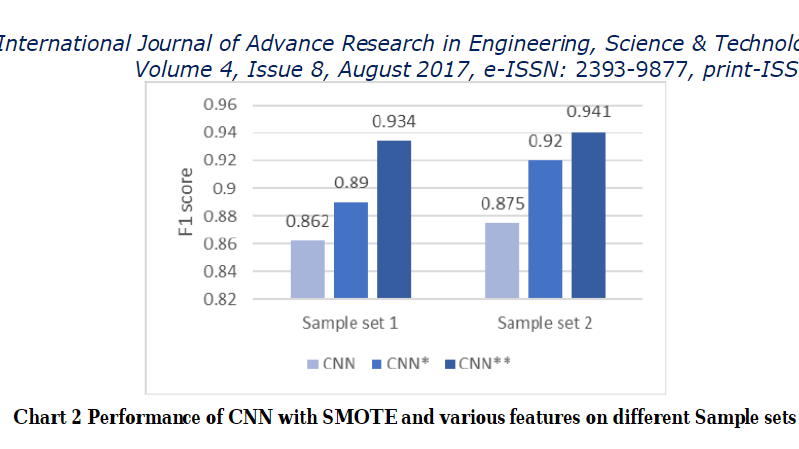
NN\*\*= NN\* with added feature

CNN\*= CNN with SMOTE

CNN\*\*= CNN\* with SMOTE







Conclusion:

In this paper, Neural network and Convolutional Neural network is applied with SMOTE.

We have also transformed features and added new feature that gives

better performance by increased number of TP and decreased number of FP.

Comparison result shows that in CNN, performance of precision is poor

than NN because ratio of legitimate transactions detected as fraudulent one is more than neural network. On other hand in CNN, ratio of detecting fraudulent transaction

is more than NN which improves performance of recall and F1 score.

Results show that CNN with SMOTE and feature

transformation overcome issue of precision and outperforms NN in all terms. Limitation is fraudulent transaction which behaviour is same as legitimate transactions can’t be detected.

references:

[1]Statista the statistic portal (2017, March 14) available https://www.statista.com/topics/871/online-shopping/

[2]Tanmay Kumar Behera, SuvasiniPanigrahi, “Credit Card Fraud Detection: A Hybrid Approach Using Fuzzy Clustering & Neural

Network”, IEEE Computer Society, 2015

[3]Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang, “Credit Card Fraud Detection Using Convolutional Neural Networks”, Springer

International Publishing AG 2016.

[4]Smt.S.Rajani, Prof.M. Padmavathamma, “A Model for Rule Based Fraud Detection in telecommunications”, International Journal of

Engineering Research & Technology (IJERT), Vol. 1 Issue 5, July – 2012.

[5]Hidden Markov model (2017, March 15) available https://en.wikipedia.org/wiki/Hidden\_Markov\_model

[6]Abhinav Srivastava, Amlan Kundu, Shamik Sural, Senior Member, IEEE, and Arun K. Majumdar, Senior Member, IEEE , “Credit Card

Fraud Detection Using Hidden Markov Model” , IEEE transactions on dependable and secure computing, vol. 5, no. 1, january-march

2008.

[7]Y. Sahin and E. Duman, “Detecting Credit Card Fraud by Decision Trees and Support Vector Machines”, IMECS vol 1, 2011.

[8]Michael Nielsen (2017, March 15), Deep learning available http://neuralnetworksanddeeplearning.com/chap6.html

[9]Ghosh, S., Reilly, D.L.: Credit card fraud detection with a neural-network. In: Proceedings of the Twenty-Seventh Hawaii International

Conference on System Sciences, 1994, vol. 3, pp. 621–630. IEEE (1994)

[10]EminAleskerov, Bernd fieisleben and Bharat Rao, “CARDWATCH: A Neural Network based database Mining System for Credit Card

Fraud Detection”

[11]Sam Maes, Karl Tuyls, Bram Vanschoenwinkel, Bernard Manderick, “Credit card fraud detection using bayesian and neural networks”,

International Naiso Congress on Neuro Fuzzy Technology, 2002.

[12]Metasyn(2016, October 5), available at https://github.com/metasyn/creditcardfrauddata

[13]Chawla, N.V., Hall, L.O., Bowyer, K.W., Kegelmeyer, W.P. (2002) “SMOTE: Synthetic Minority Oversampling Technique”, Journal of

Artificial Intelligence Research, vol.16, pp.321-357.

[14]Alejandro Correa Bahnsen, DjamilaAouada, Aleksandar Stojanovic, Björn Ottersten, “Feature engineering strategies for credit card fraud

detection”, 0957-4174/ 2016 Elsevier.

[15]Smt.S.Rajani, Prof.M. Padmavathamma, “A Model for Rule Based Fraud Detection in telecommunications”, International Journal of

Engineering Research & Technology (IJERT), Vol. 1 Issue 5, July – 2012.