Analysis on Credit Card Fraud Detection Methods

Wazid Ansari [1] and Ankita Sharma [2]

[1] B.TECH, Delhi Technological University

[2] Masters of Business Administration, A.P.J. Abdul Kalam Technical University

Chapter 1

1. Introduction

In this era of digital transformation, money transactions have become cashless, one of the most preferred ways for customers to shop offline and online is through credit cards. The use of credit cards has seen an upward trend especially in countries like India. There are several reasons consumers shift from debit cards to credit cards including cashback, reward points, EMI options, and premium memberships, etc. With that said, the fraud incidents associated with credit cards have also multiplied in recent years. Simply put, when an individual uses someone else's credit card for his/her personal advantage without the knowledge of card owners it is called credit card fraud. In addition to this, according to ACFE (The Association Of Certified Fraud Examiners), credit card fraud is defined as, "*the use of one's occupation for personal enrichment through the deliberate misuse or application of the employing organization's resources or assets*". Also, governments, financial institutions, and individuals bear huge financial losses around the globe every year due to credit card fraud and dearth of the effective fraud detection system.

There are myriads of factors on which credit card fraud are detected:

1. **Theft/Loss Fraud:** In this credit card fraud the perpetrator will use someone else's card many times until the card is blocked by the bank. The earliest the card owner will contact the concerned bank the fastest the bank will take appropriate measures to stop the transactions.

2. **Counterfeit Fraud:** In this credit card fraud the card details are remotely used by the perpetrator. They may copy the card details via any website where a physical card is not required. Often the online merchants are at risk of losing hefty amounts as they offer customer payments by credit cards.
3. **Application Fraud:** When the perpetrator applies for a credit card with false information it is known as application fraud. In addition to this, in this type of fraud, the fraudsters imitate genuine customer information to acquire a credit card.
4. **PoS Fraud:** When a small skimming device is attached to a PoS (Point Of Sale) device to steal the user's data at the time of transaction it is called PoS fraud.

In recent years many countries around the globe have suffered millions of dollars of losses to card frauds. According to NPCI 641 customers, 90 ATMs endured 13 million (1.3 Crores) Rupee losses into deceitful transactions. Annual report published by IC3 Online fraud activities have increased manifold from 2008 to 2017 due to plethora of online fraud such as identity fraud, stolen credit card, and credit card fraud. It is clear that financial losses have been

amplified, especially credit card frauds. This study pays attention to the banking sector domain, best fraud detection methods to deploy and detect fraudsters as well prevention of cyberattacks to the system.

Furthermore, in the initial phase detection of fraud is imperative to stop the deceit activity. Besides this, numerous fraud prevention and detection mechanisms have been utilized to avert and detect card frauds. Additionally, the credit card fraud prevention mechanism does not only stop the fraudulent activities but also abets in impeding the cybercrime attacks on the network systems and using prevention methods such as cryptography algorithms, PIN, firewall, internet security system, and spike detection to name a few. To detect fraud the second stage mechanism i.e. fraud detection is used and applied in varied machine learning, data mining, and bio-inspired techniques such as Support Vector Machine (SVM), Decision Tree, Artificial Neural Network, K Nearest Neighbor (KNN), Genetic Algorithm (GA), Artificial Immune System (AIS) and Bayesian Network, Logistic Regression, Fuzzy Logic, Hidden Markov Model (HMM). The core objective of these techniques is to understand the customer's patterns and fraud patterns in both normal and fraudulent transactions.

It is imperative for credit card issuers to identify the fraudulent credit card transaction so that the cardholder is not charged for the items they have not purchased. Additionally, these problems can be solved by using machine learning methods which cannot be overstated along with new-age data science techniques. In the technology-driven era credit card fraud is a very relevant problem that calls for attention from the data science and machine learning field to get a definite solution to this problem.

Chapter 2

2.1 Historical Background

One of the major ethical issues in the credit card sector is fraud. The primary objective of any financial institute is to detect the different types of credit card frauds and then implement different strategies that can be used in fraud detection. To add, the earliest credit card fraud that has happened in history was either via dumpster diving (carbon copies of credit card receipts given to the customer by the merchant) or by filching someone's credit card from the wallet. However with advancements in technology the use of the aforementioned credit card fraud techniques has decreased. Besides this, one of the easiest ways to procure genuine credit card holder details or in some cases even the credit card is via postal theft. That said, a fraudster who has nabbed the credit card has the customer's personal information including credit card limit, bank information, and card number. So, the fraudster can employ the same information to procure a new card or create a new bank account whereas the true owner is clueless about the same activity.

2.2 Some famous credit card fraud from the history

The TJX Companies, Inc (an American multinational store) had a data breach of systems between July 2005 and January 2007, where 45.6 million credit card data was exposed and the key accused was Albert Gonzalez. To date, the biggest credit card fraud that has occurred was led by Albert Gonzalez in August 2009- 130 million credit card

and debit cards were stolen at various companies in the United States, such as Heartland Payment Systems, 7-Eleven, and Hannaford Brothers.

In 2012, Adobe disclosed that the network of hackers comprised 40 million sets of crucial payment card information. The information incorporates encrypted payment card numbers, customer names, orders related information, and card expiration details.

In 2013 between November and December, at Target Corporation about 40 million credit cards breach of systems happened where imperative customer information like name, expiry date, account number, and security codes were stolen.

The usage of credit cards in developing as well as developed countries have seen an upward trend. People use credit cards for shopping, paying bills, traveling, and many more other things. However, as credit card users have increased so the credit card fraud cases have also increased. The need of the hour is to have effective credit card fraud detection techniques to tackle this major problem. According to the Nilson Report, the payment card fraud losses have reached $28.65 billion in the year 2019. These frauds affect not only consumers but merchants and issuers also. Additionally, the growth of the e-commerce sector and extensively available internet has created new opportunities for fraudsters to explore, they are using sophisticated hacking methods such as cyberlaundering, phishing, cyber extortion, and cybersquatting.

Chapter 3

Literature Review

In 2011, Siddhartha Bhattacharya et.al conducted a comprehensive comparative study of random forest and support vector machine which reveals that the most accurate information is procured by random forest technique followed by support vector machine and logistic regression. In 2015, J. Esmaily et.al have suggested a hybrid of artificial neural networks and decision trees. The two-phase approach was used in their study, in the first phase a new set of data was obtained by employing the decision tree and multilayer [ML] technique, and then data is inserted into the Multilayer perceptron to classify the data. The outcome of this proposed study was a low false detection rate. In 2011, Raghavendra Patidar has suggested an amalgamation of Neural networks and Genetic Algorithms. The neural network is used to assort the transaction whereas the genetic algorithm is employed to optimize the solution. In 2015 Tanmay Kumar et.al The proposed study followed a hybrid approach by using fuzzy clustering and neural network techniques to credit card fraud detection. The study is conducted in two-phase, in the first phase they undertake a c-means clustering algorithm to originate a suspicious score and in the second phase to determine if the transaction is suspicious or not it is entered into the neural network.

Y. Sahin et.al suggested credit card fraud detection using a hybrid of support vector machines (SVMs) and decision trees. When the data set was small the decision tree provides accurate results whereas when the data set increased SVMs provides accuracy.

Machine language because of its plethora of applications and less time consumption is the most used technology in credit card fraud detection. Machine learning is usually based on an algorithm that enables a computer to study and proceed via experience. Additionally, machine learning is applied in numerous fields namely medical, regression, banking, and diagnosis, etc. Algorithm and statistical models are combined in machine learning to allow the electronic machine to perform the activities then based on the outcome a model is constructed through data which is then examined on the trained model. In addition to this, deep learning is an integral part of machine learning that exercises neural networks. To add, Convolution neural networks, recurrent neural networks, artificial neural networks, and autoencoders are some of the most used methods that come under deep learning.

Furthermore, the Artificial Neural Network (ANN) is based on the structure and function of the human brain. An AI (artificial intelligence) function mimics the human brain working in processing data as well as creating patterns that abet in the decision-making area. To add, ANN is also known as neural networks or multilayer perceptrons [Figure 1]. These neural networks can be comprehended using backpropagation learning algorithms.

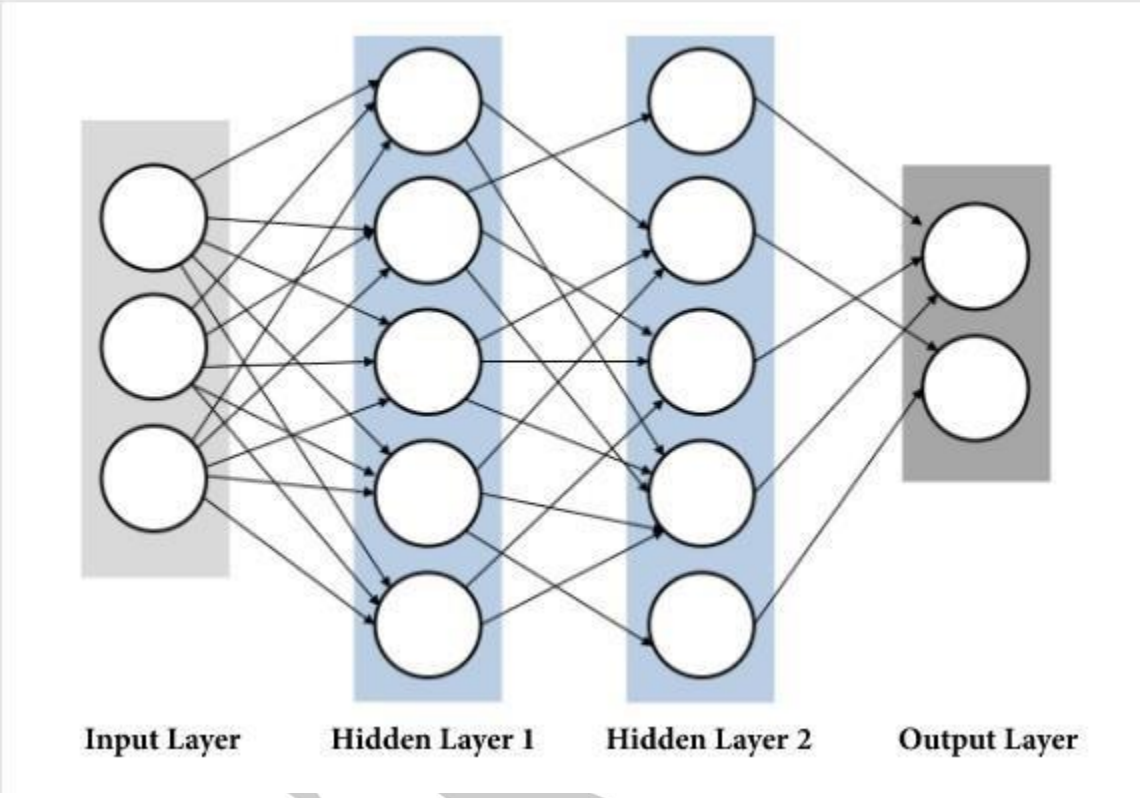

Fig. 1 Artificial Neural Network Configuration

Chapter 4

Main Objectives

In this paper, our objective is to explore hidden patterns by employing the most used credit card fraud detection techniques. That said, all fraudulent transactions undergo a similar pattern. So, by employing numerous pattern

recognition systems namely logistic regression, KNN classifier, SVM classifier, decision tree classifier, random forest classifier any financial transaction can be classified as a fraudulent transaction or not.

Chapter 5

Methodology

Right now, many classification techniques like - Hidden Markov Model, Artificial Neural Networks, Support Vector Machine (SVM), and Decision tree etc. are present to solve many industrial problems. In this paper we will use many such techniques for solving a classification problem.

5.1 Data Collection

Right now many websites are available for collecting the Data like - kaggle, Socrata, Datasets on Github, Google Public Datasets, Data.gov etc.

Even learning websites like geeksforgeeks, hackerearth, greatlearning, etc provide data for learning and practice. Here, we will use a dataset from one of these learning websites known as hackerearth.

Recently hackerearth launched a competition on Data Science where they provided data for Credit card fraud detection and we are going to use that data in this paper.

5.2 Steps involved in this paper for prediction:

1. Downloading data
2. Fixing null values
3. Splitting into training and test set
4. Feature Reduction
5. Encoding categorical feature
6. Feature Scaling
7. Modeling our data

5.3 Techniques used for classification

1. Hidden Markov Model

The name Markov indicates a change in the state with time. Hidden Markov Model (HMM) is a theoretical model that has a set of states that are hidden. However, something parallel to them can be discovered, and based on that sequence

we can anticipate the order of state changes. The spending pattern of the credit card holder is one such parameter on which HMM model is trained. Plus, any change in the transaction if it varies from the customer general profile can be classified as a fraudulent transaction.

2. Artificial Neural Network

ANN aka Artificial Neural Network consists of three types of layers namely: Input, Output, And Hidden. In this model, the data/information travels from the input network to the output layers known as forwarding Propagation. Additionally, in the ANN model, the layers are trained according to cardholder normal behavior. However, any dubious transactions are backpropagated via the ANN network and from there they are categorized as fraudulent and non-fraudulent transactions. Besides this, in the credit card fraud detection method this technique has been found as the most effective technique as a neural network doesn't need to be reconditioned.

3. Support Vector Machine

SVM alias Support Vector Machine is a supervised learning algorithm in which using a hyperplane the given data set is separated. The core objective of SVM is to discover this hyperplane. Moreover, there can be numerous hyperplanes but we have to ascertain the optimal hyperplane. The parameters nearest to the hyperplane in varied classes are called support vectors. And by using this support vector new data points are obtained.

4. Decision Tree

A decision Tree is a computational tool for prediction and categorization. A tree embraced by internal nodes, branch, and leaf nodes i.e. internal node designates a test on an attribute, branch designates an outcome of that test and lead node denotes a class label. It periodically partitions the data set by employing either breadth greedy approach or depth-first greedy approach and halts when all components have been allotted a particular class. The data is further categorized into groups for partition rule to be methodical, here a single class outweighs a group.

Besides this, the other most commonly utilized techniques for credit card fraud detection methods are Support Vector Machine (SVM), Naïve Bayes (NB), and K-Nearest Neighbor algorithms (KNN). In addition, these techniques can be used in collaboration as well as alone by using meta-learning techniques. That said, amongst all the existing techniques the most popular method used is machine learning because of its simplified implementation and exceptional predictive performance.

Chapter 6

You would probably agree to this that almost every fraudulent transaction is accomplished by embracing a similar pattern. Fortunately, the persistent advancements in technology have led to the successful emergence of several pattern recognition systems. By employing these fruitful systems, it becomes possible to classify transactions that are

fraudulent. Some of the dime-a-dozen wielded systems are mentioned below along with their working. So, keep reading to get yourself enlightened!

6.1 Artificial Neural Network

This technique fundamentally entails the collaboration of the human brain with the computational power of the machine. And as the name indicates, it utilizes neurons as the deciding sites and the edges between neurons to enumerate the endowment of each neuron in the preceding layer in the decision and result at the current neuron. Primarily, the whole concept is based on pattern recognition. It is pivotal to highlight here that the previous year's data is stuffed into the network and based on this data it figures out a new incoming transaction as a swindle or fair and square one. Additionally, its training can both be overseen or unsupervised. In the first scenario, the outcome can be anticipated before and the outcome is compared with actual training. And as far as the second option is concerned, there are no results to be compared and thus the outcome cannot be predicted beforehand.

6.2 Decision Tree

To throw light on this classification strategy; it is basically considered a computational tool. Delving a bit deeper to comprehend the working; a tree constitutes internal nodes heralding a test on an attribute. Moreover, each branch signifies an outcome of that test and each leaf node embraces a class label. It periodically partitions a dataset wielding either a depth-first greedy approach or may also employ a breadth-first greedy approach. Further, it halts as soon all the elements have been allocated a specific class. In continuation, for the successful accomplishment of the partition rule efficiently, the need of the immediate hour is the separation of data into groups where a single class is known to predominate in each group. To put it simply: the recommended or desired partition will be the one that does not see the subsets overlapping. This essentially means that they are clearly disjoint to their allowed limit.

6.3 Support Vector Machines

To apprise you with an intriguing fact; apparently, it is a supervised learning algorithm and is known to separate the existing dataset into divergent classes using a hyperplane. The primary intent of SVM is to explore this hyperplane. Whilst there might be several hyperplanes evolving but we are concerned about searching for the optimal hyperplane. To divulge an interesting part; all those points appearing nearest to the hyperplane in the different classes are known by the moniker support vectors. Further, it is these vectors that are used to foresee the classes of new data points. A neophyte incoming point is targeted on the equation of the hyperplane and then is categorized as to which class it is affiliated with based on the fact on which side of the hyperplane it falls on the vector space. The machine is fed with supervised data, the one whose results are already known. It comprehends the behavior of fraudulent transactions alongside the transactions that are authentic and then it can classify new transactions as to which class it falls in.

6.4 Logistic Regression

To duel the inconsistencies of linear regression where it furnished values advancing the limit of 1 and less than 0, the concept of logistic regression comes into action. Irrespective of the moniker it is given, LR is hailed for predicting binomial and multinomial outcomes. The ultimate goal for this approach is to estimate the values of the parameter's coefficients employing the sigmoid function. Besides, the benefits of Logistic Regression are looked up every time the name "clustering" comes into action. It is also used when a transaction is ongoing. Thus, it scrutinizes the values of its attributes and indicates whether the transaction should move ahead or not.

Chapter 7

Needless to say: credit card fraud has become one of the major perils to business establishments presently. Fraudsters employ a congregating number of techniques to commit this crime. To define credit card fraud in simpler terms, it is an individual's initiative to use another individual's credit card for personal reasons without being in the knowledge of the actual user. They simply don't have any clue about this mischievous activity. Further, the individual using the card is highly unlikely to have any connection with the actual owner of the credit card.

7.1 Role of banks in preventing such frauds

With the ever-augmenting incidents of credit card frauds, such activities need to be averted right away. Fraudsters are employing divergent techniques to be successful in their motto. While such swindlers might call themselves pretty intelligent, the banks around the globe have begun employing such strategies that can overpower fraudsters' tactics. The development of robust applications aims to ferret out potential fraud and safeguard customers' accounts. All thanks to algorithms, artificial intelligence, and biometrics, which in collaboration have helped save millions of dollars while also setting the seal that account holders are readily able to oversee their funds.

At the onset, it appeared to be a mammoth task for banks and credit unions to keep up with criminals' evolving tactics; however, the recent technological advancements have emerged to be a savior. Many modern-day techniques vow to prevent frauds effectively and luckily they are able to do so. The diligent application of such techniques will finally result in a situation that will realize a complete halt of swindling activities.

7.2 Fraud Prevention Technologies

As said earlier, the development of neophyte technologies comes as salvage in preventing fraudulent transactions. Albeit fraudsters are trying their best to somehow get successful in their initiative; however, the modern-day robust applications impede them from doing so. Many fraud detection technologies enable merchants and banks to carry out highly automated and sophisticated screenings of incoming transactions. As a result, it becomes feasible to get a breakthrough on suspicious transactions.

Have a look at some of the commonly wielded fraud prevention techniques that promise so many things:

· **Manual Review**

Any kind of evaluation that primarily involves human intervention is known by the moniker-manual review. During this procedure, every transaction is meticulously monitored to check for its authenticity. Since the brain of humans is entailed, this modality promises higher efficacy.

However, as is the case with every method, there is a disadvantage involved. The concept of manual review might boast about its impeccable potency but proves to be bling. This means a lot of money needs to be expended for hiring individuals to gauge millions of daily credit card transactions. What's more, since the use of machines is totally ditched, the time consumed for accomplishing manual transactions is also on the higher side. It also lacks to determine some of the most prevalent patterns of fraud, such as the use of a single card multiple times in several locations.

· **Address Verification System**

The Address Verification System, which is commonly abbreviated as AVS is fundamentally a tool furnished by credit card processors and issuing banks to merchants to envision suspicious credit card transactions and to obstruct the path of credit card fraud. Delving a bit deeper; the AVS cross verifies the billing address stuffed in by the card user to that of the actual address submitted initially. If there is some mismatch, this technique automatically alarms the merchant. The credit card processor is known to send a response code back to the merchant heralding the degree of address matching. In simpler words: it empowers the merchant to decide whether the particular card transaction shall be accepted or rejected. One of the known drawbacks of this system is its incompetent nature in the case of international transactions.

· **Card Verification Method**

The Card Verification Method constitutes a 3 or 4 digit numeric code, which is essentially printed on the card but at the same time not embossed. This means that it is neither available in the magnetic stripe. Therefore, when the user goes to purchase something, the merchant can request the cardholder to divulge this numeric code in case of a card-not-present transaction. The objective of CVM is to make sure that the authorization is properly done to ascertain that the user giving the details is the actual holder of the card. Albeit this modality provides some security measures to the merchant, but at the same time, it can't prevent the occurrences of transactions placed through physically stolen cards. Furthermore, fraudsters who are in temporary possession of the card can read and provide the merchant with the same code.

· **Payer Authentication**

This emerging technology has made its strong presence felt ever since it officially hit the market. It was launched worldwide by Visa way back in the year 2002 and promises to bring in a new level of security to business-to-consumer internet commerce. In this method, the user is issued a Personal Identification pin at the time they are handed a credit card. This pin is similar to the one that many banks issue to their subscribers when distributing debit cards. At the time of executing an online transaction, this pin needs to be essentially stuffed.

· **Biometrics**

It is a modality through which the quirky details of the cardholder like a fingerprint or his/her signature are recorded and read by a computer machine. Hence, the stored information can be compared during the advent of future transactions to ensure that the right person in actuality is performing the transaction.

Chapter 8

8.1 Study design

8.1.1 Data Analysis

Downloading data

First step is to import our data into the IDE. We are using python for the analysis because python has a lot of useful libraries which allow us to code the problem without explicitly coding everything.

Libraries that we are going to use are:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
```

```
df = pd.read_csv('train.csv')      # Importing Data
```

Feature exploration

We have 19 columns and 43,508 rows.

In our dataset, a good amount of columns are present. Let's take a look into these columns and check if we can assume something about these columns.

```
df.info()                          # Looking for columns
```

0 customer_id	8 no_of_days_employed	16 prev_defaults
1 name	9 occupation_type	17 default_in_last_6months
2 age	10 total_family_members	18 credit_card_default
3 gender	11 migrant_worker	
4 owns_car	12 yearly_debt_payments	
5 owns_house	13 credit_limit	
6 no_of_children	14 credit_limit_used(%)	
7 net_yearly_income	15 credit_score	

Table 1. These are the columns that we have to deal with.

We have 19 columns out of which credit_card_default is our target feature containing 0 and 1 value.

Class 0 represents valid transactions while class 1 represents fraud transactions.

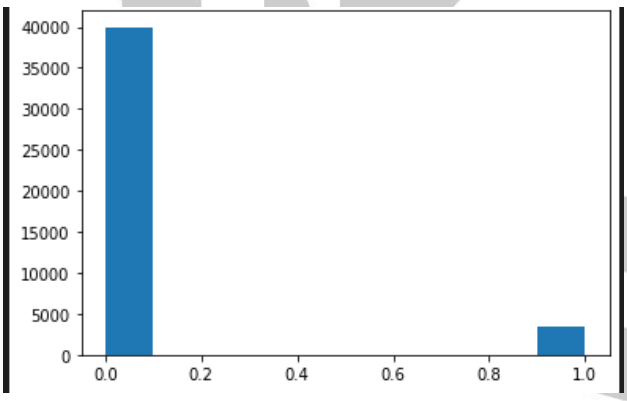

Fig. 2 Fraud Vs Valid Transactions visualization

Above figure shows fraud transactions are very less as compared to valid transactions. Therefore, the outliers are the real culprit here for doing fraud,

So we cannot just remove these outliers.

Fixing null values.

Null values can reduce the performance of our model so in order to increase the performance of our model, we have to handle these null values.

Our dataset contains null values but in very less quantity.

df.isnull().mean()*100 # Looking for percentage of missing values

Column Name	Percentage of null values
customer_id	0.00%
name	0.00%
age	0.00%
gender	0.00%
owns_car	1.20%
owns_house	0.00%
no_of_children	1.70%
net_yearly_income	0.00%
no_of_days_employed	1.02%
occupation_type	0.00%
total_family_members	0.18%
migrant_worker	0.19%
yearly_debt_payments	0.20%
credit_limit	0.00%
credit_limit_used(%)	0.00%
credit_score	0.02%
prev_defaults	0.00%
default_in_last_6months	0.00%
credit_card_default	0.00%

Table 2. Null percentage in each column.

Since our dataset doesn't contain missing values more than 2 % so it is safe to remove these rows.

df = df.dropna() # Deleting missing rows

Splitting our data into dependent and independent variable

For prediction, we need to declare a target variable and input variable, for that we are going to split our data into X and Y.

X = df.iloc[:,-1] # Independent Variable

```
Y = df.iloc[:,18]                    # Dependent Variable
```

Splitting our data into training and test set

Now we will split our data for training and testing. Our 80% of data will train and on the remaining 20% of data we will try to predict the values and calculate accuracy score for the prediction.

```
X_train,X_test,y_train,y_test = train_test_split(
                                    X,Y,test_size=0.2,
                                    random_state = 42)
```

Now that we have so many features to deal with, we can ease our work by reducing some features. Even irrelevant features can actually reduce the performance of a machine learning model. For example: KNN algorithm works of nearest neighbor logic, if it find the nearest value from the irrelevant column then our model performance will decrease. So, it's a good practice to reduce the columns.

Let's see the relation of every column with each other, for that we will use correlation heatmap.

Correlation basically tells how one feature is changing with another feature.Its value ranges from -1 to 1.If it is +1 then it means one feature is increasing with another feature and if the value is -1 then one feature is decreasing while the other one is increasing.

Feature Reduction

```
sns.heatmap(X_train.corr(),cmap='PuBu',annot=True)
```

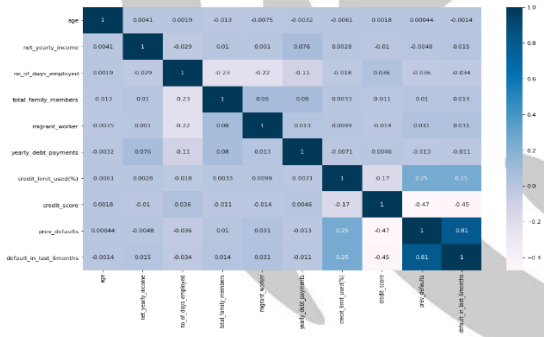

Fig. 3 Correlation heatmap

From the above matrix, the correlation value between:

- no_of_children & total_family_member > 0.8

- net_yearly_income & credit_limit > 0.8
- prev_defaults & default_in_last_6months > 0.8

(Here 0.8 is our threshold value, above which is strong positive correlation)

So it is safe to remove 3 columns from above 6 columns.

Our dataset contains 3 data type columns:

1. object
2. int64
3. float64

Handling categorical Data.

- Since a machine learning model cannot handle categorical features so we need to encode categorical features.
- Our dataset contains nominal categories features so we will use OneHotEncoder to encode our dataset.

Ohe = OneHotEncoder(sparse=False,handle_unknown='ignore)

Ohe.fit_transform(X_train,y_train)

Ohe.fit(X_test,y_test)

Scaling Data.

It is necessary to scale our data because some machine learning algorithms like Logistics regression, KNN, SVM uses distance calculation and If we don't scale our data then the bigger value will dominate the lower value by huge difference and our model will not perform properly.

sc = StandardScaler()

sc.fit(X_train,y_train)

Now our dataset is ready for training.

We will now call the algorithms

8.1.2 Modeling our Data

Since we are using python, it is easy to deploy the machine algorithm libraries without explicitly programming.

Now we are creating objects for each Algorithm class.

lr = LogisticRegression()

rf = RandomForestClassifier()

kn = KNeighborsClassifier()

svm = SVC()

dt = DecisionTreeClassifier()

Now that we have created the objects, we can fit these algorithms in our training set.

Fitting these objects using fit() function

lr.fit(X_ftrain,y_train)

rf.fit(X_ftrain,y_train)

kn.fit(X_ftrain,y_train)

svm.fit(X_ftrain,y_train)

dt.fit(X_ftrain,y_train)

Predicting test results

Now we are going to calculate the value of dependent variable:

y_pred1 = lr.predict(X_ftest)

y_pred2 = kn.predict(X_ftest)

y_pred3 = svm.predict(X_ftest)

y_pred4 = dt.predict(X_ftest)

y_pred5 = rf.predict(X_ftest)

9 Results

Calculating R2 Coefficient of determination:

Now, we are going to see how well our model is performing based on the R2 coefficient of determination.

It basically tells how our predicted values are varying from the mean of actual value.

R2 = 1 - SSR/ SSM

SSR = sum(actual - predicted)2

SSM = sum(actual - mean)2

Calculating Accuracy score of each algorithm

```
print("Logistic Regression : ",round(accuracy_score(y_pred1,y_test)*100,2))
print("KNN Classifier score : ",round(accuracy_score(y_pred2,y_test)*100,2))
print("SVM Classifier score : ",round(accuracy_score(y_pred3,y_test)*100,2))
print("Decision Classifier score : ",round(accuracy_score(y_pred4,y_test)*100,2))
print("Random Classifier score : ",round(accuracy_score(y_pred5,y_test)*100,2))
```

Logistic Regression : 95.69

KNN Classifier score : 91.01

SVM Classifier score : 91.51

Decision Classifier score : 97.13

Random Classifier score : 97.91

Chapter 10: Conclusion

Whilst the persistent progression in technology has led to the successful emergence of several fraud detection techniques; however, none is able to investigate all frauds completely. Unlike, detecting it the same moment the fraud is committed; the fingers are pointed out only once the deceit practice has been successfully accomplished. This is fundamental because a very minute number of transactions out of total ones are actually fraudulent. Hence, it makes sense to explore such a technology that has the potency to envision a fraudulent transaction the moment it is taking place rather than detecting it at a later stage. If this is achieved, the swindling activity can be halted there and then, and that too at a minimal cost.

Therefore, the stress should be on developing such a technology that is accurate, precise and could detect tampering of credit card details online. It should throw alarm bells as soon as some fraudulent activity is realized. The minute any tampering is done, an alarm should begin to pop up.

The major downside of all the prevailing techniques is that none available carries the potency to set the seal on giving the same results in all environments. Albeit there isn't much denying the fact that they can give better results with a particular type of dataset. However, when it comes to other datasets the same consistency is not seen.

Random Forest > Decision Tree > Logistic regression > SVM > KNN

The technique in the form of SVM is lauded for giving excellent results when it comes to small data sets. But the real problem is realized when large datasets are used.

Additionally, the likes of decision trees bestow better results on sampled and pre-processed data. On the flip side, when it comes to logistic regression, it is known to provide better accuracies with raw unsampled data.

Random forest on the other hand gives the best result for determining the fraudulent.

To overcome the hassles mentioned above, a hybrid type model of the existing techniques is endorsed by many maestros. It will aid in overcoming the contemporary limitations while side by side also ensuring enhanced performance. This type of model has been relentlessly put forward by several professionals. They are also of the view that in order to develop a good hybrid model, it is imperative to pair it with an expensive technique that takes a long to train but provides highly accurate and precise results. After all, the ultimate motive is to design an optimization technique to lessen the overall cost of the system and make the system train faster. Moreover, the choice of techniques for a hybrid will fundamentally rely on the applications and environment of the fraud detection system.

Chapter 11

References

Glynis D. Morris BA, FCA, ProfessorPatrick Dunne BSc, MBA, in Non-Executive Director's Handbook (Second Edition), 2008

Aleskerov, E., Freisleben, B. & B Rao. 1997. 'CARDWATCH: A Neural Network-Based Database Mining System for Credit Card Fraud Detection', Proc. of the IEEE/IAFE on Computational Intelligence for Financial Engineering, 220-226.

Chan, P., Fan, W. Prodromidis, A. & S Stolfo. 1999. 'Distributed Data Mining in Credit Card Fraud Detection'. IEEE Intelligent Systems, 14; 67-74.

Anderson, R. 2007. The Credit Scoring Toolkit: theory and practice for retail credit risk management and decision automation. New York: Oxford University Press.

Adnan M (2011) Detect CNP fraudulent transactions. World Comput Sci Inf Technol J 1(8):326–332

Dheepa V, Dhanapal R (2012) Behavior based credit card fraud detection using support vector machines. J Soft Comput 2(4):391–399

Siddhartha Bhattacharyya, Sanjeev Jha, Kurian Tharakunnel, J. Christopher Westland, Data mining for credit card fraud: A comparative study, Decision Support Systems, Volume 50, Issue 3, 2011, Pages 602-613, ISSN 0167-9236, https://doi.org/10.1016/j.dss.2010.08.008

J. O. Awoyemi, A. O. Adetunmbi and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," 2017 International Conference on Computing Networking and Informatics (ICCNI), 2017, pp. 1-9, doi: 10.1109/ICCNI.2017.8123782.

R. M. jamail esmaily, “Intrusion detection system based on multilayer perceptron neural networks and decision tree,” in International conference on Information and Knowledge Technology, 2015

Raghavendra Patidar and Lokesh Sharma, “International Journal of soft computing and engineering”, vol. 1, no. NCAI2011, 2011

Tanmay kumar behera, “credit card fraud detection: a hybrid approach using fuzzy clustering and neural network,” in international conference on advances in computing and communication Engineering, 2015

E. D. Y. Sahin, “Detecting credit card fraud by decision trees,” in Proceedings of the international multiconference of engineers and computer science, Hong Kong, 2011

Krebs, Brian (4 October 2014). "Adobe hacked: customer data, source code compromised". The Sydney Morning Herald. The Sydney Morning Herald Newspaper

A.A. Taha, S.J. Malebary , An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine, IEEE Access 8 (2020) 25579–25587, doi: 10.1109/ACCESS.2020.2971354.

P. Kumar, F. Iqbal, Credit card fraud identification using machine learning ap- proaches, in: Proceedings of the 1st International Conference on Innovations in Infor- mation and Communication Technology (ICIICT), CHENNAI, India, 2019, pp. 1–4, doi: 10.1109/ICIICT1.2019.8741490 .

Asha RB, Suresh Kumar KR, Credit card fraud detection using artificial neural network, Global Transitions Proceedings, Volume 2, Issue 1, 2021, Pages 35-41, ISSN 2666-285X, https://doi.org/10.1016/j.gltp.2021.01.006.

Dahee Choi, Kyungho Lee, "An Artificial Intelligence Approach to Financial Fraud Detection under IoT Environment: A Survey and Implementation", Security and Communication Networks, vol. 2018, Article ID 5483472, 15 pages, 2018. https://doi.org/10.1155/2018/5483472

S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, "Credit card fraud detection: a fusion approach using dempstershafer theory and bayesian learning," Information Fusion, vol. 10, no. 4, pp. 354–363, 2009.

V. Sharma et al., "Cooperative trust relaying and privacy preservation via edge-crowdsourcing in social Internet of Things," Generation Computer Systems, 2017.View at: Google Scholar

K. RamaKalyani and D. UmaDevi, "Fraud detection of credit card payment system by genetic algorithm," International Journal of Scientific & Engineering Research, vol. 3, no. 7, 2012.

Snehal patil, "credit card fraud detection using decision tree induction algorithm," international journal of computer science and mobile computing, vol. 4, no. 4, pp. 92-95

Masoumeh Zareapoor, Pourya Shamsolmoali, Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier, Procedia Computer Science, Volume 48, 2015, Pages 679-685, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.04.201

J. Kumar and A. K. Singh, "Workload prediction in cloud using artificial neural network and adaptive differential evolution," Future Generation Computer Systems, 81, 41-52, 2018.

J. Cheng and R. Greiner, "Learning bayesian belief network classifiers: Algorithms and system," In Conference of the Canadian Society for Computational Studies of Intelligence, pp. 141-151, Springer, Berlin, Heidelberg, June-2011

Y. Pandey, "Credit card fraud detection using deep learning," International Journal of Advanced Research in Computer Science, May – June 2017

Zareapoor, Masoumeh & K.R., Seeja & Alam, Afshar. (2012). Analysis on Credit Card Fraud Detection Techniques: Based on Certain Design Criteria. International Journal of Computer Applications. 52. 35-42. 10.5120/8184-1538