Grenze International Journal of Engineering and Technology, Jan Issue

Credit Card Fraud Detection using Machine Learning

Sushma V[1], Shriya Neelamma[2], Yathin machaiah[3] and Suroor Fathima[4]
[1-4]ATME College of Engineering/Computer Science Department, Mysuru
Email: sushma.sush13@gmail.com, shriyaneelamma@gmail.com, yathinmacho@gmail.com, suroorfathima80@gmail.com

***Abstract*—The amount of online transactions is growing these days to a large number. A big portion of these transactions contains credit card transactions. The growth of online fraud, on the other hand, is notable, which is generally a result of ease of access to edge technology for everyone. There has been research done on many models and methods for credit card fraud prevention and detection. Artificial Immune Systems is one of them. However, organizations need accuracy along with speed in the fraud detection systems, which is not completely gained yet. In this process we mined the transaction of each data and it finally predict the fraud detection of data.**

***Index Terms*— Cohesive soil, Ricehuskash, Metakaolin and Cement.**

I. INTRODUCTION

Credit card fraud is an important issue and has considerable cost for banks and card issuer companies. Financial organizations try to prevent account misuse using different security solutions. The more complex the security solutions are, them or sophisticated fraudsters get i.e. fraudsters change their methods overtime. Therefore it is crucial to improve fraud detection methods along with security modules which try to prevent fraud. Fraud detection has become a crucial activity in order to decrease the impact of fraudulent transactions on service delivery, costs, and reputation of the company.

There are plenty of methods used for fraud detection each of which tries to retain maximum quality of service while keeping false alarm rate at minimum. Fraud is cost and detecting it before the transaction is registered will reduce this cost significantly, which needs a very accurate system with quite few false alarms. Edge and Falcone Sampaio state that while implementation of proactive methods increases the potential for early fraud alerting, real-time processing significantly reduces the available time window within which computational analysis should be performed and an accurate decision should be made in response to newly arriving transactions.

II. LITERATURE REVIEW

A) Data mining using rules extracted from SVM: an application to churn prediction in bank credit cards.

Data mining (also called as Knowledge Discovery in Database) is a process that consists of applying data analysis and discovery algorithms that produce particular enumeration of pattern (or model) over the data [1]. Data mining has been efficiently used in wide range of profiling practices, such as manufacturing [2], fraud detection [3].Increasing number of customers has made the banks conscious of the quality of the services they offer. The problem of customers shifting loyalties from one bank to another has become common. This phenomenon, called 'churn' occurs due to reasons such as availability of latest technology, customer-friendly staff and proximity of geographical location, etc. Hence, there is a pressing need to develop models

Grenze ID:01.GIJET.9.1.619

that canpredict which existing 'loyal' customer is going to churn out in near future.Research shows that customers with longer relationships with the firm have higherprior cumulative satisfaction ratings than online bank customers. It is moreprofitable to segment and target customers on the basis of their (changing) purchase behavior and service experiences rather thanon the basis of their (stable) demograph-ics. Churn management consists of developing techniques that enable firms tokeep their profitable customers and aims at increasing customer loyalty.

B) Applying data mining to insurance customer churn management. Author: Soeini, R.A., Rodpysh, K.V.

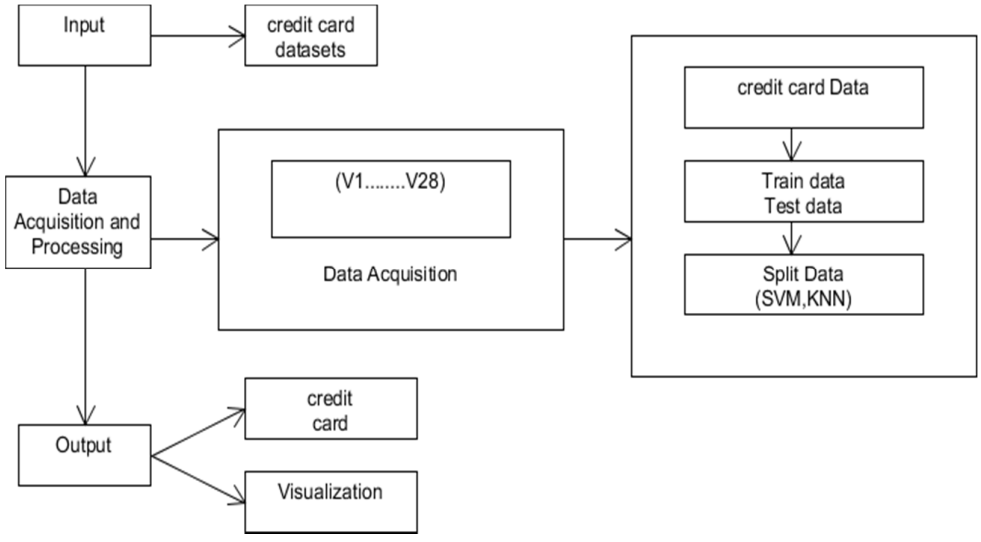
According to competition in insurance industry in Iran in recent years and entrance of private sector, keeping customers has become more important for insurer companies and reasons of churning is challenging. Thus in this research, data mining methods is used for Customer churn management (CCM). In first step, customers with equal characteristics were selected by clustering K-means method and in the second step, using churn index and decision tree CART, reasons of customer churn were analysed. Data mining process was done by Clementine software on set of data gathered from seven "Iran Insurance" branches in Anzali as population size. Costumer clustering and knowing the reasons of churning by decision tree CART make Company choose better policy to reduce that.

C) A Survey: Data Leakage Detection Techniques. Author: K. S. Wagh Objective: This Paper Analyzes latest techniques and proposed novel sampling algorithm – based data leakage detection

III. METHODOLOGY

As compared Credit card fraud events take place frequently and then result in huge financial losses. The number of online transactions has grown in large quantities and online credit card transactions hold a huge share of these transactions. Therefore, banks and financial institutions offer credit card fraud detection applications much value and demand. Fraudulent transactions can occur in various ways and can be put into different categories. System Architecture: Description: The credit card fraud detection Diagram example contains two processes, and one datasets file. Based on the diagram, we know that the one is collection of data sets and pre processing those datasets for example handling missing values, data analyses, splitting data step sand predicting the result for binary classification. Activity Diagram: Description: In this credit card fraud detection Activity Diagrams describe how activities are coordinated to provide a service which can be at different levels of abstraction. In this activity diagram we have mention the main abstraction for the credit card fraud detection. Data all stored in the csv file, after fetching the data sets we are going to apply process like data pre processing, data preprocessing, data modeling, algorithm and accuracy.

System Architecture:



IV. Results And Discussions

Dataset Model

In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

%matplotlib inline
```

```
C:\Users\admin\Anaconda3\lib\site-packages\statsmodels\tools\_testing.py:19: FutureWarning: pandas.util.testing is deprecated.
Use the functions in the public API at pandas.testing instead.
  import pandas.util.testing as tm
```

In [2]:
```
data = pd.read_csv(r'E:\WiproVIT2020\Creditt card\credit_card/creditcard.csv')
data.columns
```

Out[2]:
```
Index(['Time', 'V1', 'V2', 'V3', 'V4', 'V5', 'V6', 'V7', 'V8', 'V9', 'V10',
       'V11', 'V12', 'V13', 'V14', 'V15', 'V16', 'V17', 'V18', 'V19', 'V20',
       'V21', 'V22', 'V23', 'V24', 'V25', 'V26', 'V27', 'V28', 'Amount',
       'Class'],
      dtype='object')
```

In [7]:
```
# Check the data summary
data.describe()
```

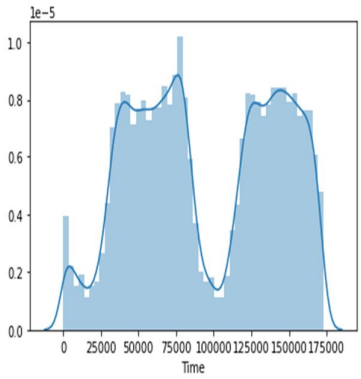
Out[7]:

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	..
count	284807.000000	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	..
mean	94813.859575	3.919560e-15	5.688174e-16	-8.769071e-15	2.782312e-15	-1.552563e-15	2.010663e-15	-1.694249e-15	-1.927028e-16	-3.137024e-15	..
std	47488.145955	1.958696e+00	1.651309e+00	1.516255e+00	1.415869e+00	1.380247e+00	1.332271e+00	1.237094e+00	1.194353e+00	1.098632e+00	..
min	0.000000	-5.640751e+01	-7.271573e+01	-4.832559e+01	-5.683171e+00	-1.137433e+02	-2.616051e+01	-4.355724e+01	-7.321672e+01	-1.343407e+01	..
25%	54201.500000	-9.203734e-01	-5.985499e-01	-8.903648e-01	-8.486401e-01	-6.915971e-01	-7.682956e-01	-5.540759e-01	-2.086297e-01	-6.430976e-01	..
50%	84692.000000	1.810880e-02	6.548556e-02	1.798463e-01	-1.984653e-02	-5.433583e-02	-2.741871e-01	4.010308e-02	2.235804e-02	-5.142873e-02	..
75%	139320.500000	1.315642e+00	8.037239e-01	1.027196e+00	7.433413e-01	6.119264e-01	3.985649e-01	5.704361e-01	3.273459e-01	5.971390e-01	..
max	172792.000000	2.454930e+00	2.205773e+01	9.382558e+00	1.687534e+01	3.480167e+01	7.330163e+01	1.205895e+02	2.000721e+01	1.559499e+01	..

8 rows × 31 columns

```
sns.distplot(data['Time'])
```

```
<AxesSubplot:xlabel='Time'>
```



V. Conclusion

Credit card fraud is without a doubt an act of criminal dishonesty. This article has listed out the most common methods of fraud along with their detection methods and reviewed recent findings in this field. This paper has also explained in detail, how machine learning can be applied to get better results in fraud detection along with the algorithm, pseudo code, explanation its implementation and experimentation results. While the algorithm does reach over 99.6%accuracy, its precision remains only at 28% when a tenth of the data set is taken in to consideration. However, when the entire data satisfied into the algorithm, the precision rises to 33%. This high percentage of accuracy is to be expected due to the huge imbalance between the number of valid and number of genuine transactions.

VI. Future Work

In future work, we intend to enhance the performance and take the security and privacy of the data in real time into consideration. As we have considered very less number of fraud datasets for the prediction, we have to consider large number of data sets so to increase accuracy result.

References

[1] A.Vellidoa,P.J.G.Lisboaa,J.Vaughan"Neuralnetworksinbusiness:asurveyofapplications".Elsevier,ExpertSystemswithApplications,(1999).17;(51–70).

[2] A.J.GraaffA.P.Engel brecht "The Artificial Immune System for Fraud Detection in the Telecommunications Environment";(2011).(1-4)

[3] Abhinav Srivastava, Amlan Kundu, Shamik Sural, Arun K. Majumdar. "Credit Card Fraud Detection using Hidden Markov Model". IEEE Transactions on dependable and secure computing ,Volume 5;(2008)(37-48).

[4] Aihua Shen Rencheng Tong,Yaochen Deng "Application of Classification Models on Credit Card Fraud Detection".(2007).

[5] Anshul Singh, Devesh Narayan"A Survey on Hidden Markov Model for Credit Card Fraud Detection". International Journal of Engineering and Advanced Technology (IJEAT), (2012). Volume-1,Issue-3; (49-52).

[6] B. Sanjaya Gandhi , R.Lalu Naik, S.Gopi Krishna, K.lakshminadh "Markova Scheme for Credit Card Fraud Detection". International Conference on Advanced Computing, Communication and Networks;(2011).(144-147).

[7] Bidgoli, B. M., Kashy, D., Kortemeyer, G. & Punch, W. F "Predicting student performance: An Application of data mining methods with the educational web-based system LON-CAPA". In Proceedings of ASEE/IEEE frontiersin education conference. .(2003).

[8] Bolton,R.J.,Hand,D.J(2002)."Statistical fraud detection: A review".Statistical Science(1994).28(3);(235—255).

[9] Clifton Phua,Vincent Lee, Kate Smith,and Ross Gayler"A comprehensive survey of data mining-based fraud detection research".In Artificial Intelligence Review .(2005).

[10] Cortes,C.&Vapnik,V"Supportvectornetworks,MachineLearning"..(1995).Vol.20;(273–297).