Journal Homepage: - www.journalijar.com
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)
Article DOI: 10.21474/IJAR01/4582
DOI URL: http://dx.doi.org/10.21474/IJAR01/4582
ISSN NO. 2320-5407

RESEARCH ARTICLE

MACHINE LEARNING BASED CREDIT CARD FRAUD ANALYSIS, MODELING, DETECTION AND DEPLOYMENT.

Shivkumar Goel[1] and Hitesh Patil[2].

1. Deputy H.O.D, Dept. of MCA, VESIT, Mumbai, Maharashtra, India.
2. P.G Student, Dept. of MCA, VESIT, Mumbai, Maharashtra, India.

Manuscript Info

Manuscript History

Received: 23 April 2017
Final Accepted: 25 May 2017
Published: June 2017

Key words:-
Credit Card Fraud Analysis, SCQ, Apache, ANN, Clustering.

Abstract

Credit card fraud is critical business risk that every stakeholder of financial system including issuer, acquirer etc. has to manage tightly to ensure business continuity and credibility of payment system. As the popularity of the credit card payment as one of the online payment mode is increasing more and more due to the revolution that has taken place in E-commerce sector. Traditional fraud models designed years back deliver near about 70% accuracy and were meeting business needs till this time. However fraudsters are increasing gaming the system to create new types of frauds which has resulted in consistent decrease in model accuracy. The fraudulent transactions and real transactions are scattered all around and there is very little difference to distinguish between them. Many techniques based on Artificial Intelligence, Data mining, Fuzzy logic, Sequence Alignment, Genetic Programming, Machine learning has evolved in detecting various credit card fraudulent transactions. This paper represents how the combinations of different clustering and machine learning algorithm which can best adapt to the changing scenarios taking place can be used and deployed on a very large scale to detect the fraudulent transactions and use to ensure the credibility of the payment system.

Introduction:-

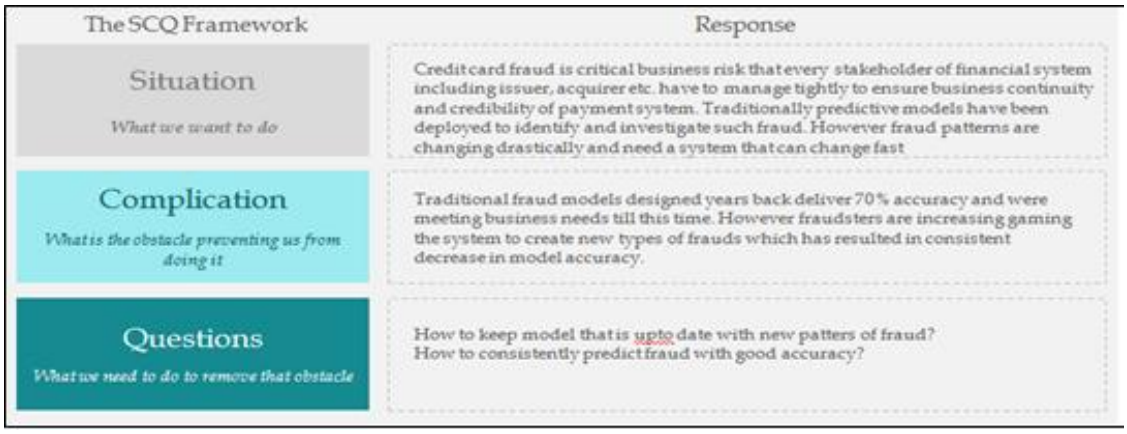
Credit card fraud can be defined as "Unauthorized account activity by a person for which the account was not intended. Operationally, this is an event for which action can be taken to stop the abuse in progress and incorporate risk management practices to protect against similar actions in the future". In simple terms, Credit Card Fraud is defined as when an individual uses another individuals credit card for personal reasons while the owner of the card and the card issuer are not aware of the fact that the card is being used. And the persons using the card has not at all having the connection with the cardholder or the issuer and has no intention of making the repayments for the purchase they done[3].

So we can use data analytics to tackle fraud as there are lot of weaknesses in the internal controlling systems. In order to effectively test and monitor internal controls, we need to look at every transaction that takes place and test them against established parameters, across applications, across systems, from dissimilar applications and data sources[1].

Corresponding Author:- Hitesh Patil.
Address:- P.G. Student, Dept. of MCA, VESIT, Mumbai, Maharashtra, India.

Most internal control systems simply cannot handle this. Due to this the organization suffers by financial losses and individual users are too affected as their information of credit card get steal. So it is important to find a solution which classifies a transaction into fraud or non-fraud. Many techniques have been developed for credit card fraud detection like Artificial Intelligence & Machine Learning and also based on locations [4].

In this paper, we focus on Machine Learning technique, basically it provides a system which is supposed to classify a current transaction into fraud or non-fraud. In this paper, we are taking credit card fraud detection problem as a classification problem. Many classification algorithm have been developed [5], but the most popular one is Decision Tree. Basically there are two technique for credit card fraud detection: 1. Supervised 2. Un-supervised. These are the machine learning techniques, in which the first one uses training data, to build the model, which have all the attributes including class label i.e. it already contains the attribute which tells whether this previous transaction is fraud or not. And in the second technique, training data does not contain the class label i.e. this technique is class less. More study on these can be found in [5]. This paper mainly tells us about our approach towards credit card fraud analysis model building that was done by using SCQ analysis technique to factor mapping and model building and deployment of the model.



SCQ Framework.

Methodology:-

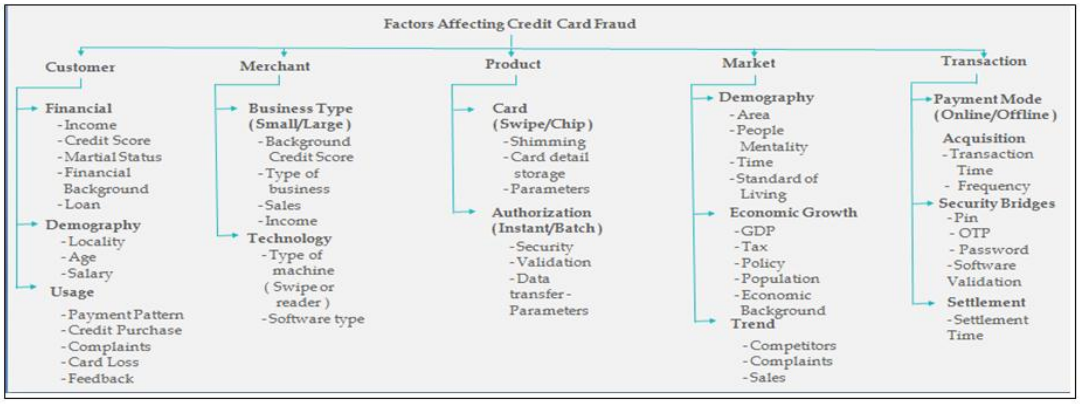
Defining the problem properly is half problem solved. Framing every question for proper definition of hypothesis is required. Barbara principle provided us with a methodology that assisted us in guiding our thinking process and helped us to develop hypothesis that helped us to solve the problem in more realistic way. Barabara Minto's Pyramid principle is a hierarchically structured thinking and communication technique that can was used to precede good structured writing[2].The core of Minto's thinking method is to group Ideas to the presenter thought process into small clusters that support the main thesis in increasing the granularity[2]. The SCQ framework helped to move through a path with insights to practical outcomes. The SCQ framework is framework that divides the problem statement as follows:

Situation- where are we now? : This helps in establishing relevance. It tells us what we exactly want to do.
Complication: It tells us: What is the obstacle preventing us from doing it?
Questions: What we need to do to remove that obstacle?
THE SCQ framework helped us to define the hypothesis which helped in model building that was used in predicting the fraudulent transactions.

Factor Mapping:-

Factor mapping is a technique that is being used to connect different factors that are mainly associated with credit cards. It help us to know who actually uses credit card, what impact it may have demographically, what are types of transactions that involve use of credit card , which types of business involve more usage of it etc. It helps in better understanding and analysis of the data, user's habit. The data analysis received by factor mapping can provide various data input to different marketing strategies, help in increasing the creditability of the economic strategies, help in developing different credit plan as per the use of the customer.

We tried to find and list down different factors that affect the credit card fraud. After listing down the different factors we distributed them across different sectors. Firstly the Problem statement was divided into bigger clusters and then each cluster analysis was carried out to divide the cluster into smaller granular clusters. After dividing the problem into smaller granular structure it was found that each granular factor was independent of each other and the impact factor was different on the credit card fraud.



Factor Mapping Tree.

Out factor mapping for fraud analysis had five main factors as: 1. Customer, 2. Merchant, 3.Product, 4. Market and 5. Transactions.

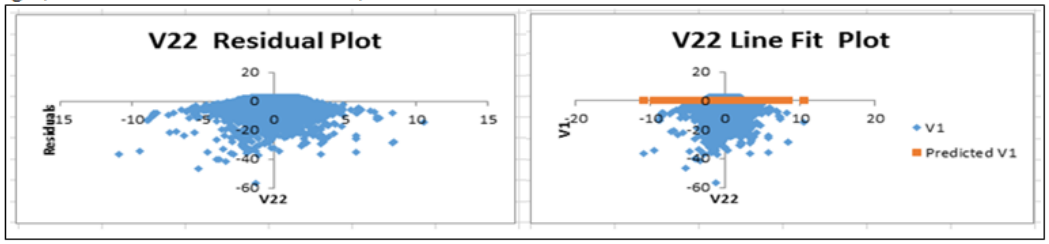
After finding the main factors then detail analysis was carried out on each factor. In case of Customer, who are the main users and the initialisers of the credit card Customers were divided into 3 main categories based upon their financial status, demographic conditions and their usage patterns. Based on the data received the detail analysis was carried out on personal traits considering his financial Background, Loan, Credit score, Marital status, his locality, age group background. Then different patterns in his transactions were carried out on his credit purchases, feedback received etc.

The factor merchant was categorized based upon his business type (Small or Big) and type of technology he use for swiping the credit card. Based on the business types, sales and income and the pattern the merchants were marked for their genuineness.

The Product factor depends upon the fact of type of card being used: the type of card, the manner in which all the parameters are stored and the manner in which all the necessary parameters were passed, the type of security, the type of validations used in each card etc were analyzed.

The market factor depended upon Demography economic growth of that particular region, the trends of the competitors. Each area based upon the demography was thoroughly analyzed. Based upon the historical data thee area, the population of that particular area, the standard of living of that area, the tax policy of that region, GDP effect.

The Transaction patterns the most important factor of all consist of payment mode (Online /Offline). Security bridges, Settlements and settlement time, Validation of the software etc.



Initial Analysis of Attributes

All the factors that can have an impact on the credit card were listed down. Based upon the thorough analysis of the factors and the data available EDA i.e. Exploratory data analysis was carried out and then the granular factors were used to define the hypothesis that can lead to fraud were found.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	65535
R Square	-1.51292E-14
Adjusted R Square	-3.51117E-06
Standard Error	1.958699243
Observations	284807

ANOVA

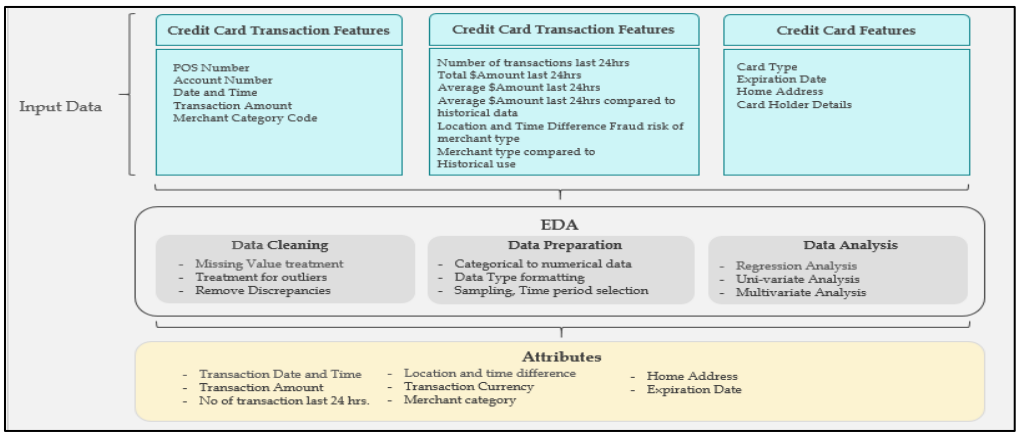
	df	SS	MS	F	Significance F
Regression	1	-1.6531E-08	-1.6531E-08	-4.30887E-09	#NUM!
Residual	284805	1092655.158	3.836502723		
Total	284806	1092655.158			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	3.85478E-15	0.003670223	1.05029E-12	1	-0.007193535	0.007193535	-0.007193535	0.007193535
V22	9.60334E-16	0.005057491	1.89884E-13	1	-0.009912543	0.009912543	-0.009912543	0.009912543

Initial Analysis of Attributes.

EDA:-

The credit card has many parameters. The parameters were categorized as three types: The transactions features, Historical Transaction features and card transaction features. The data from all the three categories were the input to the system. The data that was collected and then scrutinized and then treatment was given to the data. The treatment to the data consists of Data Cleaning, Data Preparation and data analysis.



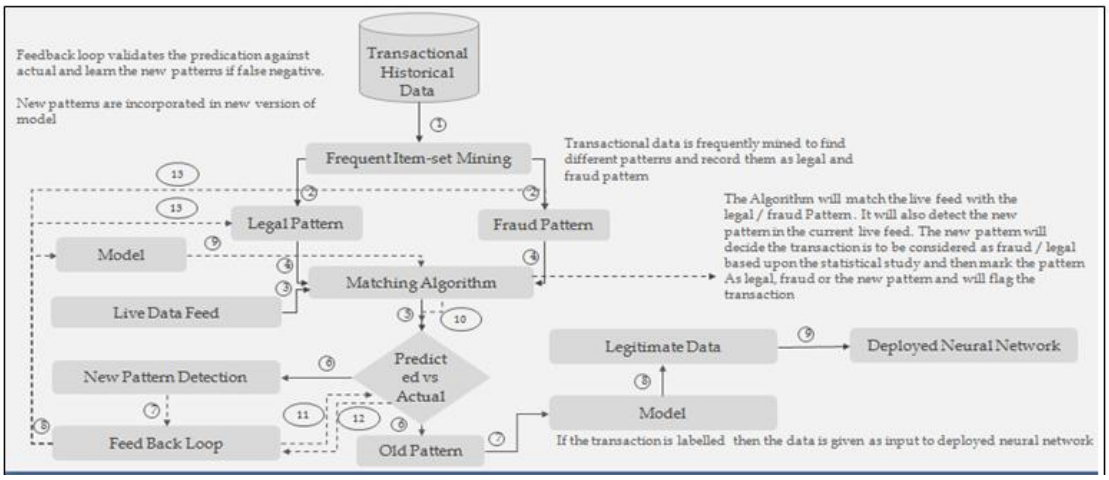
Detail EDA Procedure.

The Data Cleaning consists of treatment to the missing value, the handling of the outliers and removal of any discrepancies. The data preparation consists of type of data the categorical data or continuous data. The data analysis consists of Regression Analysis, Univariate analysis, Multivariate analysis etc.

After carrying out the statistical analysis of the data the attributes of the data that will be used for further analysis were decided and the parameters that will be given as the actual input to the model were decided.

Model Development:-

Traditionally for detecting the fraud many data analysis techniques were used but the changing technical world couldn't rely on this techniques as they cannot have an impact on the credibility of the stakeholders. Every model required to be more versatile to resist in this changing technical exploration going all round the globe. There are many techniques methods like Knowledge Discovery in Databases (KDD), Data Mining, Machine Learning and statistics available all around but they commonly fall around two categories: Statistical Techniques and Artificial Intelligence.



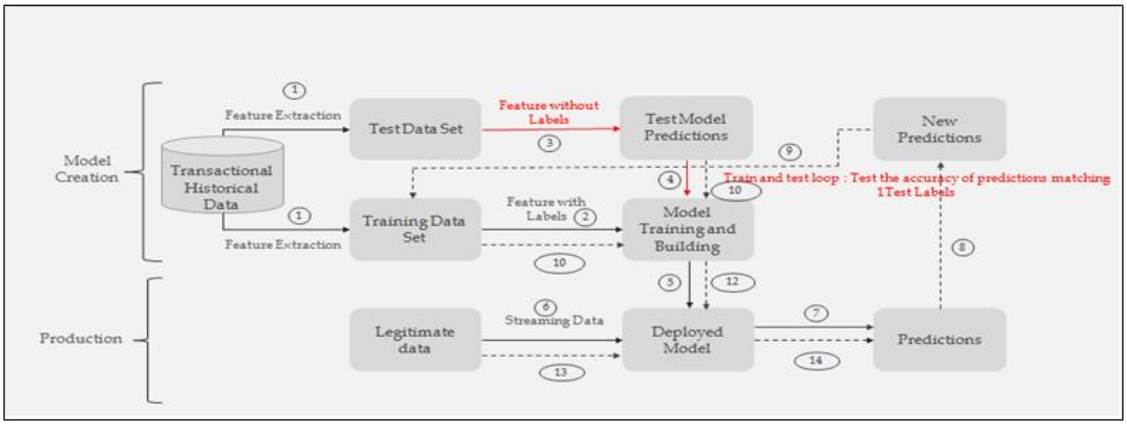
Pattern Matching Model Using Clustering Algorithm.

Our model is divided into two parts one is used to find the patterns using clustering algorithm and then pass the data through artificial neural network that will self-learn and will be robust enough to detect any types of fraud. This model will help us not only to detect the new fraud but also provide a means to analyze data.

The model will use data mining techniques to mine the data and find different fraudulent transaction patterns. This data mining technique will be used to frequently mine the data set so all the patterns are always up to dated. After the data mining the patterns that are formed will be divided into two groups Fraud Pattern and Legal Patterns. They are nothing but the two clusters.

There will be a matching algorithm whose main task is to match the pattern with the real time data that is feed to the system. The main task of this matching algorithm is to also detect the new patterns as well as to find any discrepancies

Once the pattern is matched with the existing pattern and if the transaction is true then it is labeled as legitimate data that will be further passed as an input to Neural network that will analysis and increase the accuracy of fraudulent detection system. But if the transaction detected as Fraud then it is labeled as fraud and then the transaction is denied. But if any new pattern is detected then that data stored can a feedback loop is started which will add the data to the transactional data stored and will again mine the system for data pattern and then it will be giving as an input to matching algorithm and will check for the accuracy level if still the pattern cannot be decided then there can be annual intervention triggered or the transaction can be marked as fraud and then provide it as an input to neural network that will more accurately will be able to termed the data as fraud or legal by adjusting the weights and make the system more robust.



Prediction Model Using ANN.

The input from the above model is given to the Artificial Neural Network that prevents fraudulent transactions from taking place. The Artificial Neural network model consist of input nodes who provide their respective output to intermediate node and finally the output from the intermediate node is given to the output node where actual output is compared with the expected output with certain accuracy set. While the model is tuned to get the output of exact accuracy level that is usually done by adjusting the weight at each nodes by using back propagation technique. The model is being build using tensor flow library.

Result and Insights:-

The data analysis was carried out and the model was tested for its accuracy. After the model was build it was found that the model could accurately predict the fraudulent transaction and decline the transaction with an accuracy level of approximately 99%. The relationship and accuracy level could be seen from the below result. The data was analyzed and was tried to search the patterns when actually the fraud was taking

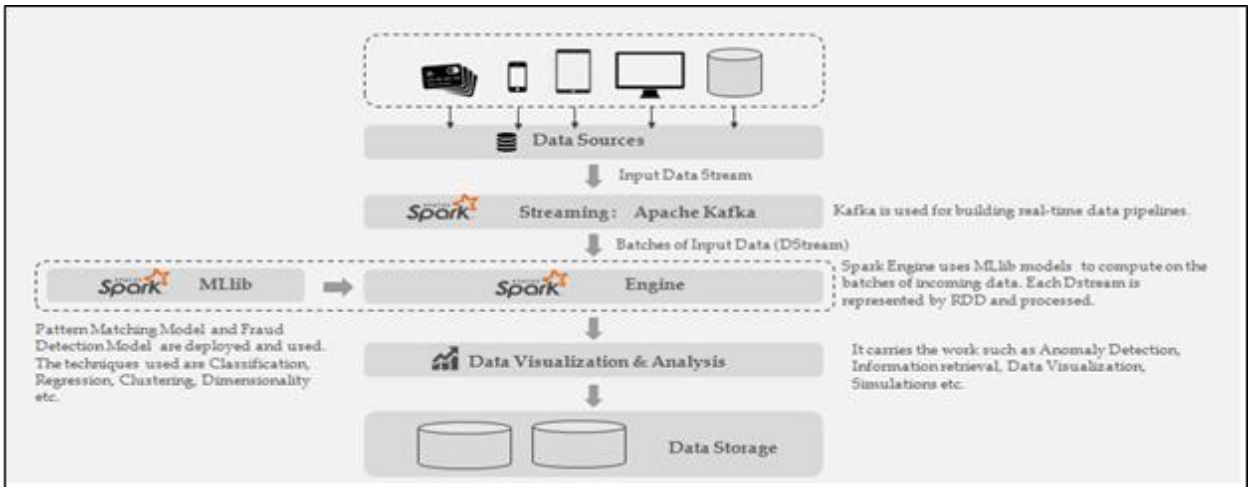
Result and Insights of Transaction Analysis.

Place. After analyzing the transaction it was found that there was not specific period where whole during the day where fraud can occurred fraudulent transaction were scattered whole during the day. So it very difficult to predict the exact time when fraud can take place. After the analysis was done there was one more outcome that fraud usually took place for smaller amount transaction. The transaction amount was smaller but the transaction frequency and occurrence of the transaction was different. The fraudulent and legal transaction visualization is shown using T-SNE visualization tool.

Deployment Architecture:-

After the model is being built most important question is always asked is how the model is going to be deployed and can be brought to use. The below given model explains how exactly the machine learning algorithm can be deployed. Credit card can be used from different sources it can be swipe transaction or an online transaction. The

card parameters has to go through this model that was build earlier and provide an instant feedback of the status of the transaction ie it is approved or rejected.The processing has to be taken place quickly because there are millllions of transaction taking place within seconds. and it has to be done accurately otherwise it will result into an hampering of the crediability of the system. The parameters of each transaction are pipelined using A pache Kafka. The apache Kafka streamlines the process into batches.



Deployment Architecture.

So that all the transaction can process into batch instead of one after the other. The output of the Kafka is given to the actual Apache engine which uses MLlib where actually our model is deployed. MLlib consist of our two models One for pattern matching and second the ANN model . The outcome of the model need to be visualized and analyses so that can be done using different data visualization and Analytical tool. This provides us with the graphical representation and visualization techniques.

Conclusion:-

Credit card fraud has become rampant in recent years. To improve the proper risk management level in an automatic and effective way, building an accurate and easy handling credit card risk monitoring system is one of the key tasks. One aim of this study is to identify the user model that best identifies fraud cases and building a model that will be highly robust to detect any change in pattern and take corrective steps by itself in order to minimize the losses and the risks involved. This paper gives contribution towards the effective ways of credit card fraudulent detection analysis and building a model along with its deployment procedure.

References:-

1. https://www.acl.com/pdfs/ACL_fraud_ebook.pdf
2. http://www.12manage.com/methods_minto_pyramid_principle.html
3. https://www.ijsce.org/attachments/File/NCAI2011/IJSCE_NCAI2011_025.pdf
4. Nadeem Akhtar, Farid ul Haq, " Real Time Online Banking Fraud Detection Using Loaction Information",International Conference on Computational Intelligence and Information Technology – CIIT 2011, Pune, India
5. K. Cios, W. Pedrycs, and R. Swiniarski, Data Mining Methods for Knowledge Discovery. Boston: Kluwer Academic Publishers, 1998.
6. Renu and SumantAnalysis on Credit Card Fraud Detection Methods, http://www.ijcttjournal.org/Volume8/number-1/IJCTT-V8P109.pdf
7. Linda Delamaire (UK), Hussein Abdou (UK), John Pointon (UK), "Credit card fraud and detection techniques: a review", Banks and Bank Systems, Volume 4, Issue 2, 2009 .
8. Khyati Chaudhary, Jyoti Yadav, Bhawna Mallick, "A review of Fraud Detection Techniques: Credit Card", International Journal of Computer Applications (0975 – 8887) Volume 45– No.1, May 2012 .
9. Vladimir Zaslavsky and Anna Strizhak," credit card fraud detection using selforganizing maps", information & security. An International Journal, Vol.18,2006.
10. L. Mukhanov, "Using bayesian belief networks for credit card fraud detection," in Proc. of the IASTED International conference on Artificial Intelligence and Applications, Insbruck, Austria, Feb. 2008, pp. 221– 225.
11. https://kafka.apache.org/ .