Hybrid Methods for Credit Card Fraud Detection Using K-means Clustering with Hidden Markov Model and Multilayer Perceptron Algorithm

Article · January 2016

DOI: 10.9734/BJAST/2016/21603

CITATIONS
2

READS
649

4 authors, including:

Stephen Fashoto
University of Swaziland Kwaluseni Swaziland
24 PUBLICATIONS **19** CITATIONS
SEE PROFILE

Olumide Owolabi
University of Abuja
20 PUBLICATIONS **112** CITATIONS
SEE PROFILE

British Journal of Applied Science & Technology
13(5): 1-11, 2016, Article no.BJAST.21603
ISSN: 2231-0843, NLM ID: 101664541

SCIENCEDOMAIN *international*
www.sciencedomain.org

Hybrid Methods for Credit Card Fraud Detection Using K-means Clustering with Hidden Markov Model and Multilayer Perceptron Algorithm

Stephen Gbenga Fashoto[1*], Olumide Owolabi[2], Oluwafunmito Adeleye[3] and Joshua Wandera[1]

[1]Kampala International University, Kampala, Uganda.
[2]University of Abuja, Abuja, Nigeria.
[3]Redeemer's University, Ede, Osun State, Nigeria.

Authors' contributions

This work was carried out in collaboration between all authors. Author SGF is the project coordinator and shared the overall responsibility for the work with author OO, which was done in two parts. The first part on introduction and review of the related literature was carried out by authors SGF and OA while the second part on methodology and implementation was carried out by authors OO and JW. Finally, authors SGF and OO concluded the work and drafted the abstract. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJAST/2016/21603
Editor(s):
(1) Hui Li, School of Economics and Management, Zhejiang Normal University, China.
Reviewers:
(1) M. Bhanu Sridhar, GVP College of Engineering for Women, Andhra Pradesh, India.
(2) Ele, Bassey Igbo, University of Calabar, Nigeria.
Complete Peer review History: http://sciencedomain.org/review-history/12615

Review Article

Received 25th August 2015
Accepted 20th September 2015
Published 10th December 2015

ABSTRACT

The use of credit cards is fast becoming the most efficient and stress-free way of purchasing goods and services; as it can be used both physically and online. Hence, it has become imperative that we find a solution to the problem of credit card information security and also a method to detect fraudulent credit card transactions. Over the years, a number of Data Mining techniques have been applied in the area of credit card fraud detection. The focus of this paper is to model a fraud detection system that would attempt to maximally detect credit card fraud by generating clusters and analyzing the clusters generated by the dataset for anomalies. The major objective of this study is to compare the performance of two hybrid approaches in terms of the detection accuracy.

**Corresponding author: E-mail: gbengafash@yahoo.com;*

We employed hybrid methods using the K-means Clustering algorithm with Multilayer Perceptron (MLP) and the Hidden Markov Model (HMM) for this study. Our tests revealed that the detection accuracy of “MLP with K-means Clustering” is higher than the “HMM with K-means Clustering” for 80% percentage split but the reverse is the case when the “MLP with K-means Clustering” is compared with the “HMM with K-means Clustering” for 10 fold cross-validation but the accuracy is the same in the two hybrid methods for percentage split of 66%. More extensive testing with much larger datasets is however required to validate theses results.

Keywords: Credit card; credit card fraud; fraud detection; data mining; K-means clustering; HMM, MLP.

1. INTRODUCTION

In this cashless era, the most universal means by which goods and services are paid for is the use of credit card. Statistics have shown that 75% of the people in America use credit cards [1]. Credit Card Fraud is described as a situation where a person uses a credit card belonging to another person for personal motives without the permission or awareness of the card-owner [2]. The Credit Card is a plastic card issued to number of users as a mode of payment [3]. There are several issues associated with online credit card use. One of the most important is fraud, which can be carried out by both individuals and merchants. A major problem is the lack of security which could lead to credit card numbers in online databases being compromised [4]. As Nigeria is gradually going cashless, credit card usage is increasing rapidly, this is because of its ease of use in online payments and the feature “buy now pay later”

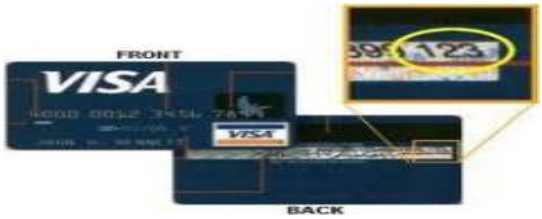

Fig. 1. A diagram of a typical credit card

There are two main types of credit cards used for making purchases:

i) Physical card

Here, there is a physical presentation of the card, for payment of goods bought or services rendered, by the card holder to the merchant. Fraudulent activities can be carried out in this type of purchase; in fact it is relatively easy, as the attacker only needs to steal the card. This could lead to significant financial loss to the credit card company if the theft is not realized quickly.

ii) Virtual card

This type of purchase is made using vital details about the credit card such as the card number, expiration date, secure code and Card Verification Value (CVV) number. Such purchases are normally done on the Internet or over the telephone. To commit fraud in these types of purchases, a fraudster needs only to know the card details [2].

Even as this rate is rapidly increasing, the means and ways by which individuals try to defraud and steal from other people is also increasing. Most fraudsters target credit card because a lot of money can be made within a short period of time and also without too many risks. Another “upside” of this fraud is that the actual crime is discovered many days after it occurs [5].

According to a global survey carried out by ACI WorldWide, in 2014, many countries have experienced relatively high fraud rates; including India, China, Germany and Sweden, to mention a few, having 41 percent, 42 percent, 16 percent and 10 percent respectively [6].

Credit card fraud increases at an almost directly proportional rate as the number of daily online and physical card users [7]. Nowadays, retailers deal more with online than regular purchases, most of which are credit card based transactions.

2. REVIEW OF RELATED LITERATURE

Most research on credit card fraud detection has focused on pattern matching in which abnormal patterns are identified from the normality. Several techniques for the detection of credit card fraud have been proposed in the last few years, some of them are briefly reviewed below.

2.1 Data Mining Application in Credit Card Fraud Detection System

[8] in her work "Data Mining Application in Credit Card Fraud Detection System" showed how an application of artificial neural networks which has built-in learning abilities, can be used to determine fraudulent and legitimate models from a huge transaction data. In this work, the Self-Organizing Map Neural Network (an un-supervised method of artificial intelligence) was used to classify credit card transactions under four clusters: low, high, risky and high-risk. Once a transaction was legitimate it was processed but if a transaction fell into any of the above clusters it was labeled fraudulent/suspicious. The fraudulent transaction will not be processed but will be committed to the database. The database interface is the entry point through which the transactions are read into the system. It is the system's interface with the banking software. In the Credit Card Fraud (CCF) detection subsystem, each transaction entering the system is passed to the host server, where the corresponding transaction profile is further checked using neural networks and general business rule. A set of data on transactions both legitimate and fraudulent (training set) is fed to the detection system. This becomes a means and criteria for identifying suspicious transactions (if they are similar to already known fraudulent transactions) which could lead to fraud (a deviation from the usual pattern of an entity usually implies the existence of fraud).

2.1.1 Credit card fraud detection using Bayesian and neural networks

Bayesian and Neural network approach has been applied as an automatic credit card fraud detection system approach and a type of artificial intelligence programming which was based on a variety of methods including machine learning approach and supervised data mining for reasoning under uncertainty [9]. Bayesian network is an acyclic graphical representation of the dependence between two variables and gives a compact specification of the joint probability distribution. The type of Bayesian network used was the Bayesian Belief Network. The type of neural network used was the Feed Forward Multi-layer Perceptron. It had three layers: The input layer, the hidden layer and the output layer.

Two learning algorithms were used, one for the Feed Forward Multi-layer Perceptron and the other for the Bayesian Belief Network (BNN).

- Back Propagation of Error Signals: This algorithm was used in learning in the Feed Forward Multilayer Perceptron. This algorithm uses two passes, the forward pass (this calculates the weighted linear combination of all its inputs) and the backward pass (this calculates the error at the output layer with respect to the desired output value for a particular pattern).
- STAGE Algorithm: This was used in learning in the BNN. It is a type of global optimization approach, which is used to identify the structure of the network. Global optimization is the problem

3. HOW HMM CAN BE USED FOR CREDIT CARD FRAUD DETECTION

After the HMM parameters are learned, the symbols from a cardholder's training data is taken and an initial sequence of symbols is formed [10].

Let $C_1,C_2,C_3, \ldots, C_K$ be one of such sequence of length K. This recorded sequence is formed from the cardholder's transactions up to time t. This sequence is imputed into the HMM and the probability of acceptance is computed. Let the probability be α_1, which can be written as

$$\alpha_1 = P(C_1,C_2,C_3, \ldots, C_K \mid \lambda) \quad (1)$$

Let C_{K+1} be the symbol generated by a new transaction at time t+1. To form another sequence of length K, we remove C_1 and add C_{K+1} in that sequence, generating $C_2,C_3, ..C_R,C_{R+1}$ as the new sequence to the HMM, we read this sequence into the HMM and calculate the probability of acceptance by the HMM. Let the new probability be α_2

$$\alpha_2 = P(C_2,C_3,C_4, \ldots ,C_{K+1} \mid \lambda) \quad (2)$$

$$\text{Let } \Delta\alpha = \alpha_1 - \alpha_2 \quad (3)$$

Assuming $\Delta\alpha > 0$, it means that the new sequence is accepted by the HMM with low probability, and it could be a fraud. The newly added transaction is determined to be fraudulent if the percentage change in the probability is above a threshold, otherwise the transaction is genuine [10,11].

$$\Delta\alpha/\alpha_1 \geq \text{Threshold} \quad (4)$$

3.1 Advantages of HMM Approach

- The detection of the fraudulent use of a card is found much faster than when using the existing system.
- In case of the existing system even the original card holder is also checked for fraud detection. But in this system no need to check the original user as we maintain a log.
- The log which is maintained will also be a proof for the bank for the transaction made.
- It is the most accurate technique for fraud detection.
- There is a decrease in the number of false positive transactions recognized as malicious by a fraud detection system even though they are really genuine [12,13].

3.2 Impact of Credit Card Fraud

The impact of credit card fraud can be explained in relation to three different persons, the cardholder (owner of the card), the merchant and the bank

3.2.1 Impact on cardholders

It is interesting to note that cardholders are the least impacted party due to the fact that consumer liability is limited by the legislation prevailing in most countries. This is true for both card-present as well as card-not- present scenarios. The consumer liability is minimized to a great extent as a result of the standards operational in many banks. Also most of the card/account holder's losses are covered, based on the cardholder protection policy. All the cardholder needs to do on seeing suspicious transactions charged to his card, is to report such charges. Based on this report, the issuing bank investigates, alongside the merchant's bank (acquirer), all information pertaining to the report, including the transaction, the cardholder's claim and the merchant's claim. After this is done, the issuing bank processes a charge back for the amount in the claim.

3.2.2 Impact on merchants

The most affected person in a credit card fraud is the merchant. Especially with transactions where the physical card is not present, the merchant bears the full loss of the fraud. The issuing bank sends a charge back to the merchant through his bank, telling him that there has been a report of suspicious charges from a legitimate cardholder; asking that the credit for the transaction be reversed. Most time it is difficult for the merchant to request a reversal of the charge back, especially if he has no physical evidence to challenge the cardholder's claim. Consequently, the cost of the fraudulent transaction is totally absorbed by the merchant.

3.2.3 Impact on banks (Issuer/Acquirer)

Sometimes the issuer or the acquirer may bear some of the costs of the fraudulent transaction, based on the scheme rules defined by both MasterCard and VisaCard. And even if the banks do not bear the loss directly, they still have to bear some indirect losses caused by the fraud. Such inevitable losses include manpower and administrative costs [13].

4. PROBLEM STATEMENT

Technology has integrated nations and the world has become a global village; the electronic market is opened to all and sundry, including thieves, hackers, and criminals. Hence, intense research has been on-going to find the most effective and efficient ways of credit card fraud detection. Different approaches such as machine learning and data mining have been used, but to train the classifiers they require labeled data for both genuine and fraudulent transactions. As mentioned earlier, it is quite difficult to get real life data for a fraud detection system; hence a system that is not totally dependent on labeled data is needed.

The key difficulties in building such a system are [14]:

1. Skewed distribution of legitimate and fraudulent data in the database that challenges the detection approaches
2. The labeled data that is meant to be used for training purpose is not readily available; as financial companies don't share their data for a number of (competitive and legal) reasons.
3. The databases that companies maintain on transaction behavior are huge and growing rapidly, which demand scalable machine learning systems.
4. Users' behavior changes quite often for different types of users, hence it is tough to track.
5. Real-time analysis is highly desirable to update models when new events are detected.

6. Easy distribution of models in a networked environment is essential to maintain up to date detection capability

5. METHODOLOGY

We employ hybrid methods using the "K-means Clustering algorithm with MLP" and "K-means Clustering algorithm with HMM" for this study. The K-means Clustering technique is selected because the credit card transactions dataset used for the study does not have a characteristic indicating whether a transaction was a suspicious fraudulent transaction or suspicious non-fraudulent transaction. So, the clustering technique is used to group credit card transactions that are suspected to be fraudulent, into a similar cluster. In the first stage of this study, we use clustering to partition the data set into distinct groups, which gives an indication of the suspicious fraudulent transactions. The output of this stage is then used to train the HMM and the MLP respectively, which can then be used to classify incoming transactions.

6. FRAUD DETECTION TECHNIQUES

Fraud Detection is the process of identifying potential or tangible scam in an organization, group or business. Fraud detection systems depend greatly on the execution of appropriate processes to identify warning signals for fraud [15]. Credit Card Fraud Detection is the process of recognizing fraudulent transactions using the general classification of transactions into both legitimate and fraudulent transactions [8]. Fraud detection techniques are divided into two: techniques based on statistical evaluation and techniques based on artificial intelligence [16].

Examples of analysis using statistical evaluation include:

- Editing incorrect data, filling up missing data, error correction, validation and detection using data processing techniques.
- Calculating parameters including performance metrics, average, probability distributions, quantiles, etc.

Fraud detection techniques that involve the use of artificial intelligence:

To segment, cluster and classify data, data mining is used; it can also find rules and associations that indicate the presence of interesting patterns, including fraud related patterns.

- Data mining uses pattern recognition for detecting clusters, suspicious behavioral patterns and approximate classes.
- Machine learning techniques that identify the characteristics of fraud automatically.
- Neural networks that can learn suspicious patterns

6.1 Credit Card Fraud Types

The ways through which fraudsters can execute a credit card fraud are numerous. The way fraudsters carry out fraudulent activities and the technology they use changes as often as technology evolves.

There are three main categories of credit card fraud; they are traditional fraud, merchant fraud and internet fraud [17,2].

6.1.1 Traditional fraud

The different methods of committing traditional credit card frauds are described below:

1) Lost/ Stolen Cards: When a card received by a legitimate account holder is misplaced or used by another person for criminal activities, it can be referred to as lost or stolen card fraud. It is one of the most difficult types of traditional fraud to deal with.
2) Account Takeover: This is when a person's valid account and personal information is legally or illegally acquired by a fraudster, who then takes over the legitimate account by providing information he/she gained; which may include the person's account number or card number. The card issuer is then contacted by the fraudster, who is now pretending to be the original account holder; he asks that the mailing address be changed to a new one; he then goes a step further by making a report of "his" lost card and requests a replacement.
3) Fake and Counterfeit Cards: There is a major threat posed by the making of counterfeit cards alongside lost / stolen cards. Fraudsters are always finding newer and more innovative ways with which to create counterfeit cards.

6.1.2 Merchant frauds

Sometimes, employees of merchant establishments or/and their employers may conspire to defraud their customers; this type of fraud is called merchant fraud. There are various ways this type of fraud can be carried out, they include:

1) Merchant Collusion: A situation where employees or their employers use customers' personal or account information to scheme a fraud against them. The fraud is not directly committed by the staff, since they pass on the information gained to fraudsters.
2) Triangulation: The fraudster in this type of fraud operates from a web site. Goods are offered at heavily discounted rates and are also shipped before payment. The fraudulent site appears to be a legitimate auction or a traditional sales site. The customer while placing orders online provides information such as name, address and valid credit card details to the site. Once fraudsters receive these details, they order goods from a legitimate site using stolen credit card information. The fraudster then purchases other goods using the credit card details of the customer. This process is designed to cause a great deal of initial confusion, and the fraudulent internet company in this manner can operate long enough to accumulate vast amount of goods purchased with stolen credit card information.

6.1.3 Internet frauds

The Internet has provided an ideal ground for fraudsters to commit credit card fraud in an easy manner. With the expansion of trans-border or "global" social, economic and political spaces, the internet has become a new world market, capturing consumers from most countries around the world. The most commonly used techniques in internet fraud are described below:

1) Site Cloning: Site cloning occurs when fraudsters clone an entire site or just the pages on which a customer places his order. Customers have no reason to believe they are not dealing with the company that they wished to purchase goods or services from because the pages that they are viewing are identical to those of the real site. The cloned or spoofed site will receive the details pertaining to the transaction and send the customer a receipt via email just as the real company would have done. The consumer suspects nothing, whilst the fraudsters have all the details they need to commit credit card fraud.
2) False Merchant Sites: These sites often offer the customer an extremely cheap service. The site requests a customer's complete credit card details such as name and address in return for access to the content of the site. Most of these sites claim to be free, but require a valid credit card number to verify an individual's age. These sites are set up to accumulate as many credit card numbers as possible. Though the sites themselves do not charge individuals for the services they provide, they are usually part of a larger criminal network that either uses the details it collects to generate profits or sell these valid credit card details to small fraudsters.
3) Credit Card Generators: Credit card number generators are computer programs that generate valid credit card numbers and expiry dates. These programs work by generating lists of credit card account numbers from a single account number using the mathematical Luhn algorithm that card issuers use to generate legitimate card number combinations. The generators allow users to illegally generate as many numbers as the user desires, in the form of any of the credit card formats, whether it be American Express, Visa or MasterCard [2].

7. HIDDEN MARKOV MODEL

A Hidden Markov Model (HMM) is a finite set of states; where a probability distribution is linked to each state. A set of probabilities called the transition probabilities direct the transition among these states. In each state an outcome can be generated, this outcome is an associated symbol of observation of the probability distribution. The only visible outcome to the external observer is the outcome at the current state, while all other states are "hidden"; hence the name "Hidden Markov Model" [12].

A Hidden Markov Model (HMM) can be considered a generalization of a mixture model where the hidden variables (or latent variables), which control the mixture component to be

selected for each observation, are related through a Markov process rather than independent of each other [18].

Mathematically, an HMM is described as having the following characteristics:

1. N is the number of states in the model. We can denote the set of state as S = $\{S_1, S_2, \ldots, S_n\}$. The state at time instant t is denoted by qt.
2. M is the number of distinct observation symbols per state. The observation symbols correspond to the physical output of the system being modeled.
3. The state transition probability matrix A= $[A_{ij}]$.
4. The observation symbol probability matrix B = $[B_{jk}]$.
5. The observation sequence O = O_1, O_2…. O_n
6. N is the number of hidden states.

It is apparent that a complete specification of an HMM requires the estimation of two model parameters, N and M, and three probability distributions A, B, and p. We use the notation (A, B, p) to indicate the complete set of parameters of model, where A, B implicitly include N & M [19].

Three different kinds of purchases are shown above they are represented as states of HMM, TT (Travel Ticket), MT (Movie Ticket), BP (Book Purchase). V and NV are two observation symbols either one is active for particular state; they are shown on each state. V indicates VIOLATION which means if incoming transaction violates the behavior sequence then V will be the observed symbol to that state and OTP (One Time Password) is sent to the customers mobile number. NV indicates NON-VIOLATION which means there is no anomaly and incoming transaction is normal [18].

Initially the normal behavior of the cardholder is used to train the HMM, then spending patterns of users can be determined using K-means clustering algorithm. Any incoming transaction that is not accepted by the HMM can be determined as suspicious. For further confirmation, a security question module that contains some personal questions that are expected to only be known to the authorized cardholder will be activated and if the transaction is fraudulent then a verification code is requested [20]. Hidden Markov model works on the Markov chain property in which the probability of each subsequent state depends on the previous state, which consists of observation probabilities, transition probabilities and initial probabilities. HMMs are commonly applied to pattern recognition tasks since they allow a formal representation of a stochastic dynamic process, and allow for a systematic analysis of the data and prediction based on such models.

8. MULTILAYER PERCEPTRON

A Multilayer Perceptron is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate output. It is a modification of the standard linear perceptron in that it uses three or more layers of neurons (nodes) with nonlinear activation functions, and is more powerful than the perceptron in that it can distinguish data that is not linearly separable, or separable by a hyper-plane [21].

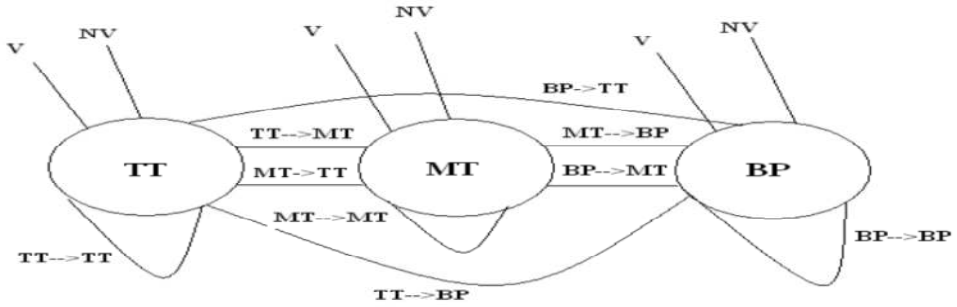

Fig. 2. An Illustration of a Hidden Markov Model

9. DESCRIPTION OF THE PROPOSED SYSTEM

9.1 The Algorithm for the Fraud Detection System (FDS)

Step 1: Log in
Step 2: Purchase what you want
Step 3: Enter your card details for verification
Step 4: The fraud check system in the bank is activated and the transaction is tested using "K-means clustering with HMM" and "K-means Clustering with MLP"
Step 5: if it is suspicious the security question module is activated
Step 6: Security question is asked.
Step 7: If the question is answered correctly, the transaction is allowed, else it is denied
Step 8: If the transaction is not suspicious it is allowed.

10. CLASSIFICATION MODELLING USING WEKA

Once the clustering model is developed, the next step of this study is developing the predictive model using the classification techniques. Since the developed clustering model does not classify new instances of the data into a certain segment, the classification process is carried out.

For starting the classification modelling experiments, the HMM and MLP algorithm is selected. The training of the HMM and MLP classification models of the experimentation is done by employing the use of training sets with 10-folds cross-validation and the percentage split classification models (using 66% and 80%).

The classification is analyzed to measure the accuracy of the classifiers in categorizing the credit card data into fraudulent and non-fraudulent classes. Accuracy refers to the percentage of correct predictions made by the model when compared with the actual classifications [22]. The classification accuracy of each of these models is reported and their performance is compared in classifying new instances of records. A separate test dataset is used for testing the performance of the classification models.

10.1 Setting Test Options in WEKA

Before we run the classification algorithm, we need to set test options. The available test options are:

1. **Use training set.** Evaluates the classifier on how well it predicts the class of the instances it was trained on.
2. **Supplied test set.** Evaluates the classifier on how well it predicts the class of a set of instances loaded from a file. Clicking on the 'Set…' button brings up a dialog allowing you to choose the file to test on.
3. **Cross-validation.** Evaluates the classifier by cross-validation, using the number of folds that are entered in the 'Folds' text field.
4. **Percentage split.** Evaluates the classifier on how well it predicts a certain percentage of the data, which is held out for testing. The amount of data held out depends on the value entered in the '%' field [23].

10.2 Confusion Matrix

The accuracy of the HMM models can be examined by confusion matrix produced by them for optimum decision on cost/benefit analysis in Table 1.

Table 1. An illustration of a confusion matrix

Actual class		Positive	Negative
	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

We employed four performance measures: precision, recall, F-measure and ROC (Receiver Operating Characteristics) space [7]. A distinguished confusion matrix (sometimes called contingency table) is obtained to calculate the four measures. A confusion matrix is a matrix representation of the classification results. It contains information about actual and predicted classifications done by a classification system. The matrix has four cells denoting, respectively, the number of samples correctly classified as true (i.e., TP - saying a Fraudulent is Fraudulent), the number of samples correctly classified as false (i.e., TN - saying Non-Fraudulent is Non-Fraudulent), the number of samples incorrectly classified as false (i.e., FN - saying a Fraudulent is Non-Fraudulent), and the number of samples incorrectly classified as true (i.e., FP - saying that Non-Fraudulent is a Fraudulent). Once the

confusion matrix is constructed, the precision, recall, F-measure are easily calculated as:

$$\text{Recall} = TP/(TP+FN) \quad (5)$$

$$\text{Precision} = TP/(TP+FP) \quad (6)$$

$$\text{F-measure} = (2*TP)/(2*TP+FP+FN) \quad (7)$$

11. FINDINGS AND DISCUSSION

Generally, from the result of the three experiments conducted on classification model of k-means clustering with HMM and classification model of k-means clustering with MLP shown in Tables 2 and 3, it could be seen that the model developed with the k-means clustering and MLP using percentage splits of 66% and 80%, the percentage split of 80% gives a better classification accuracy for predicting whether newly arriving credit card transactions are fraudulent or not. Therefore, among the different classification models built in the foregoing experimentations, the model with the percentage split of 80% in table 2 has been chosen due to its better overall classification accuracy of 51.5%.

Table 2. Overall result on classification Model of MLP with K-means clustering (hybrid approach)

Experiment	Accuracy (%)
Experiment 1 (percentage split 66%)	42.9
Experiment 2 (percentage split 80%)	51.5
Experiment 3 (10 fold cross-validation)	46.4

Table 3. Overall result on classification model with HMM and K-means Clustering (hybrid approach)

Experiment	Accuracy (%)
Experiment 1 (Percentage split 66%)	42.9
Experiment 2 (Percentage split 80%)	42.7
Experiment 3 (10 fold cross-validation)	43.9

The obtained results show that the MLP with K-Means Clustering model has better performance as compared to HMM with K-Means Clustering model.

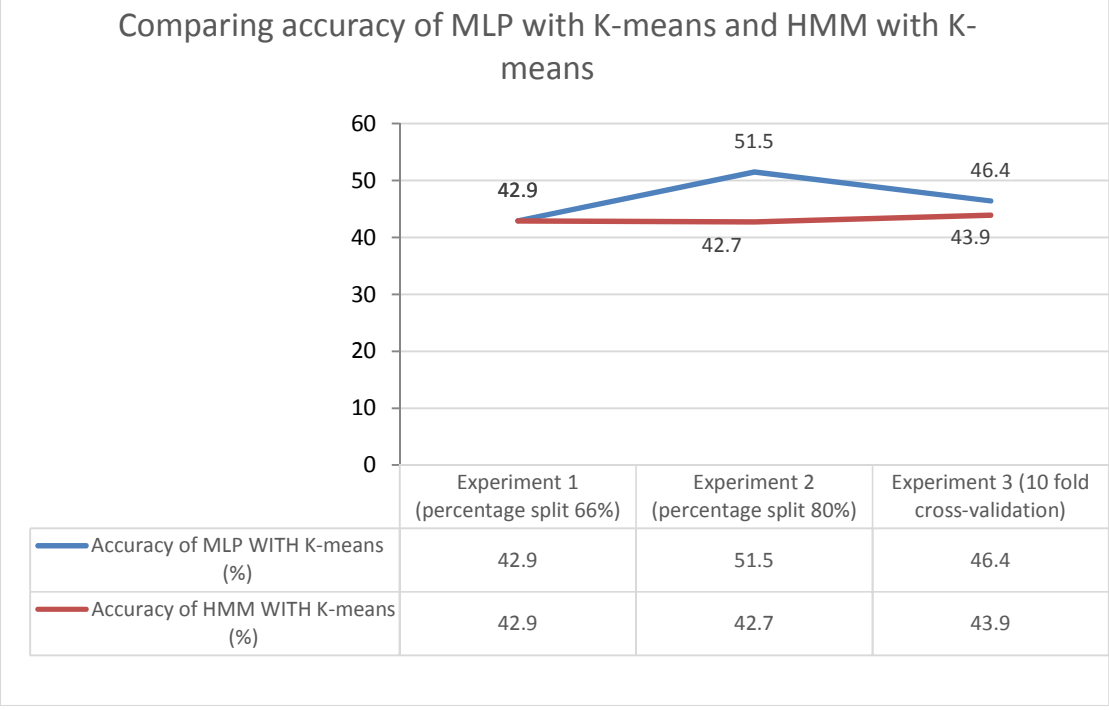

Fig. 3. Comparing accuracy of HMM With K-means and MLP with K-means

12. CONCLUSION

The study was conducted in two modules, first the clustering followed by classification model module. The initial data collected from the UCI repository did not incorporate the target class for this study. The clustering module was conducted using a self-programmed clustering algorithm and WEKA for segmenting the data into the target classes of suspicious fraudulent and suspicious non-fraudulent transactions for the Australian credit data from the UCI repository. Generally, from the result of the three experiments conducted on classification model, we can conclude as follows:

The model developed using 80% percentage split for MLP with K-means clustering performed better than the percentage split of 66% and 10-fold cross validation test option for training the classification accuracy of HMM with K-means Clustering for classifying new credit card datasets as fraudulent and non-fraudulent transactions.

The performance of K-means Clustering with MLP and HMM in detecting fraud for credit card transactions using different learning algorithms was examined in this study. The obtained results show that the “MLP with K-Means Clustering” outperformed the “HMM with K-means Clustering”.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Avinash I, Thool RC. Credit card fraud detection using Hidden Markov Model and its performance. International Journal of Advanced Research in Computer Science and Software Engineering. 2013;3(6).
2. Baeza-Yates R, Ribeiro-Neto B. Modern information retrieval. New York: ACM Press New York; 1999.
3. Binitie AP, Blamah NV, Ogah US. Synthetic software method: panacea for combating internet fraud in Nigeria. The International Journal of Engineering and Science. 2013;2(10):43-50.
4. Chandra R, Chaudhary K, Kumar A. The combination and comparison of neural networks with decision trees for wine classification. School of Science and Technology; 2007.
5. Don L, Dhakwan S. Credit card fraud detection using Hidden Markov Model. European Journal of Industrial and System Engineering. 2013;2668-3253.
6. Dunn JE. Credit and debit card fraud eating away at consumer confidence in providers.
   Available:bankITasia.com
   Available:http://bankitasia.com/bankitasia/cards/credit-and-debit-card-fraud-eating-away-at-consumer-confidence-in-providers/
   (2014, June 26)
7. Esakkiraj S, Chidambaram S. A predictive approach for fraud detection using Hidden Markov Model. International Journal of Engineering Research & Technology. 2013;1-7.
8. Fraud Facts: An Introduction To Fraud Detection. Fraud Advisory Panels; 2011
   Available:https://www.fraudadvisorypanel.org/pdf_show_157.pdf
   (Retrieved November 27, 2013)
9. Gade V, Chaudhari S. Credit card fraud detection using Hidden Markov Model. International Journal of Emerging Technology and Advanced Engineering. 2012;511-513.
10. Ingole A, Thool R. Credit card fraud detection using Hidden Markov Model and its performance. International Journal of Advanced Research in Computer Science and Software Engineering. 2013;3(6):626-632.
11. Jadhav SN, Bhandari K. Anomaly detection using Hidden Markov Model. International Journal of Computational Engineering Research. 2013;3(7):28-35.
12. Kavita R, Jyoti H. Credit card fraud detection using Hidden Markov Model. International Journal of Latest Research in Science and Technology. 2012;1(4):420-422.
13. Khyati C, Jyoti Y, Bhawna M. A review of fraud detection techniques: Credit card. International Journal of Computer Applications. 2012;45(1):39-44.
14. Kirkby R. WEKA explorer user guide for Version 3-3-4. University of Waikato; 2002.
15. Ogweleka FN. Data mining application in credit card fraud detection. Journal of Engineering, Science and Technology. 2011;6(3):311-322.
16. Phua C, Lee V, Smith K, Ross G. A comprehensive survey of data mining-

based fraud detection research. Available:https://www.google.com.ng/#q=Techniques+used+for+fraud+detection+are+categorized+in+to+two%3A+statistical+techniques+and+artificial+intelligence.+ (Retrieved November 27, 2013)
17. Raghavendra P, Lokesh S. Credit card fraud detection using neural network. International Journal of Soft Computing and Engineering. 2011;32-38.
18. Ray DP, Ghahremani Y. Credit card statistics, industry facts, debt statistics. Available:CreditCard.com Available:http://www.creditcards.com/credit-card-news/credit-card-industry-facts-personal-debt-statistics-1276.php (Retrieved October 28, 2013)
19. Srivastava A, Kundu A, Sural S, Majumdar AK. Credit card fraud detection using Hidden Markov Model. IEEE Transactions on Dependable and Secure Computing. 2008;37-48.
20. Stolfo SJ, Fan DW, Lee W, Prodromidis AL, Chan PK. Credit card fraud detection using meta-learning: Issues and initial results. New York; 1997.
21. Tripathi KK, Pavaskar AM. Survey on credit card fraud detection methods. International Journal of Emerging Technology and Advanced Engineering. 2012;*2*(11):721-726.
22. Vaibhav G, Sonal C. Credit card fraud detection using Hidden Markov Model. International Journal of Emerging Technology and Advanced Engineering. 2012;2(7):511-513.
23. Witten I, Frank E. Data mining: Practical machine learning tools and techniques. In M. Kaufmann (Ed.); 2005.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/12615