International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 24th Nov 2022. Link :http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=Issue 11

DOI: 10.48047/IJIEMR/V11/ISSUE 11/13

Title DETECTION OF CREDIT CARD FRAUDUSING MACHINE LEARNING ALGORITHMS

Volume 11, ISSUE 11, Pages: 79-87

Paper Authors
Dr. V.Uma Rani , Adaneya Sharath Chandra

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

DETECTION OF CREDIT CARD FRAUDUSING MACHINE LEARNING ALGORITHMS

Dr. V.Uma Rani , Professor, MTech, P.h.D, Department of CSE
Adaneya Sharath Chandra, MTech, Data Sciences

ABSTRACT: Credit cards have grown more popular as a method of payment for both inside and online transactions thanks to the widespread use of current electronic business systems and communications technologies. This has, however, resulted in a significant increase in credit card fraud. Many dollars are lost annually by businesses and individuals as a result of credit card fraud and thieves are always looking hunt for new methods to commit this crime. The ability to detect fraudulent transactions is a fundamental hurdle in the adoption of electronic payment. As a result, techniques for detecting credit card fraud need to be both efficient and reliable. In this paper, the authors describe an intelligent approach based on a refined light gradation boosting machine for detecting credit card fraud (OLightGBM). The proposed method involves using a Bayesian-based hyper - parameter optimization technique to fine-tune the settings of a light gradient enhancing machine (LightGBM). To demonstrate the effectiveness of our OLightGBM in spotting fraud in card transactions, we conducted experiments on two publicly accessible credit card payment data sets that included both fraudulent and genuine transactions. Comparisons using both datasets showed that the proposed approach outperformed the competitors in terms overall accuracy (98.40%), AUC (92.88%), precision (97.34%), and F1-score (0.91). (56.7 percent).

We'll be using phrases like "credit card fraud," "machine learning," "accuracy," "area under the curve," "precision," and "F1-score."

I. INTRODUCTION:

This Thing We're Ding With more and more people making purchases online and conducting financial transactions without using cash, it is more important than ever to be able to see red flags for fraud. Credit card fraud occurs when a thief makes transactions using the other user's credit card information without the cardholder's permission or knowledge. Every year, credit card fraud costs companies and customers billions of dollars due to the widespread use of credit card payments and the lack of sufficient security procedures. A reliable assessment of the loss is elusive, however, since credit card firms are generally reluctant to make such data public. However, there is some accessible data on the monetary losses associated with credit card theft. Several billion dollars are lost every year due to unsafe credit card transactions. Credit card fraud cost businesses and consumers $22.8 billion in 2017, and experts anticipate that number to climb to $31 billion by 2020. The most common forms of card fraud include both fraudulent applications and fraudulent usage of cards. Fraudulent credit card applications are one example of application fraud. When an impostor submits a request for a new card transaction using false identification information, the issuer has committed credit card fraud. Behavior fraud refers to the use of a credit card for illegitimate purposes after it has been granted legally. Credit card fraud detection has long been a difficult problem for both cardholders and financial institutions. Because of the enormous quantities of money at risk if even a small percentage of fraud cases go undetected, fraud with credit cards is also a big topic for the academic community.

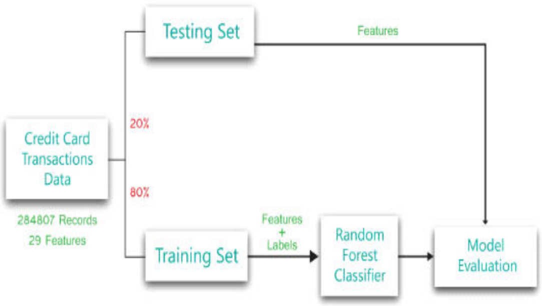


Fig.1: Example figure

- **LITERATURE REVIEW**

- **HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture,**

Credit card fraud costs card issuers billions of dollars every year. A sophisticated fraud detection system built on a state-of-the-art fraud detection model is necessary for lowering fraud losses. An improved feature engineering approach based on homogeneity-oriented behaviour analysis is the cornerstone of our learned in the classroom fraud detection system, which represents our most significant contribution (HOBA). We use a real-world dataset procured from one of the leading commercial banks in China to conduct a comparative study that would allow us to assess the performance of the proposed framework. The results of our trials demonstrate that the method we propose is an effective and realistic way to detect credit card fraud. When comparing the false positive rate of several approaches for detecting fraudulent transactions, our proposed strategy performs better than the state-of-the-art solutions. Our method has substantial managerial ramifications since it allows issuers of credit cards to use the proposed strategy, so better safeguarding their customers' interests, reducing fraud losses, and decreasing regulatory fees.

2.2 A method for online retailers to detect fraudulent credit card purchases via the use of data mining

Credit card fraud costs online stores billions of dollars every year. With the development of machine learning, researchers have identified increasingly sophisticated methods to detect fraud; nevertheless, specifics are seldom made public. In this article, we describe the process we went through to develop and launch a system for fraud detection for a Fortune 500 e-commerce company. The feasibility of merging human and automatic classification is explored, and a number of machine learning approaches are compared and contrasted. As a result, this research may help scholars and professionals create data mining based methods for identifying fraud and associated difficulties. All of the hard work that went into this project has paid off in the shape of an automated system and fresh knowledge that fraud analysts may use to improve their methods of human revision.

Adaptive deep learning techniques for identifying credit card fraud 2.3.

Credit card fraud may effect a small percentage of transactions, but the associated losses might still be significant. Since fraudsters employ a wide variety of strategies, there is an urgent need to develop automatic Detecting Fraud Systems (FDS) which can reliably spot such activities. In fact, fraud tendencies may vary greatly among nations, customer subsets, and payment platforms . It is becoming more important for transactional firms to reuse existing pipelines and adapt them for use in different domains and situations, basically the age-old difficulty of transfer learning, as the cost of establishing data-driven FDSs rises.

2.4.1 Isolation forest and local outlier factor for identifying credit card fraud

Modern technical progress is accelerating at an exponential pace, and this progress can be put to either good or evil ends. As a result of advancements in associated technologies, e-commerce and other types of online transaction have thrived, with credit cards being the preferred method of payment. The use of a credit card allows the buyer to make a purchase quickly without having to fork over the whole price in cash. Customers may make purchases at any participating retailer across the globe without ever having to pull out their wallets. While the use of plastic payment methods has increased, so too has the incidence of credit card theft. Credit card fraud is easy to pull off because of how the system works. Thieves are always on the lookout for new methods to commit credit card fraud, which is a huge issue that costs companies and consumers a significant amount of cash every year. Detecting fraudulent internet transactions is a major difficulty for banks and other financial organisations. In order to reduce losses from credit card fraud transactions, it is crucial for financial institutions such as banks to have reliable fraud detection systems in place. Many researchers have proposed various strategies for detecting and reducing such fraud. Here, we offer comprehensive experimental data and compare the Language implementations of the Outlier Factors and also the Isolation Factor approaches. We compared the accuracy of Local Outlier Factor (LOF) and the Isolated Forest (IF), and found that the LOF was 97% accurate while the IF was 76% accurate.

2.5 Identity Theft Suspicion Scoring in True Credit Applications Based on Online Communities

This research presents a speedy technique, dubbed communal analysis fear scoring (CASS), for generating numerical suspicion ratings on streamed credit application areas on implicit links between them over time and space. CASS has a number of features, such as couple communal scoring of identifying attributes for applications, definition of suspiciousness categories for implementation, incorporation of spatial and temporal weights, as well as smoothed k-wise scoring of many linked application-pairs. Data mining a large number of real credit applications reveals that CASS reduces false alarm rates while maintaining respectable hit rates. CASS is scalable for such a large dataset and may immediately detect precursors of identity theft. As an added bonus, new insights have been gleaned through studying the interdependencies of various software packages.

The Application of Support Vector Machine (svm to the Analysis of Credit Risk in Corporate Financial Institutions:

A credit risk index has been developed using actual data from commercial banks. The term "index" may refer to both financial and non indexes. In this analysis, the SVM method is used for judgement. The plan calls for selecting training sets with rising proportions. The accuracy of the percentages is determined on the size of the sample used. The model's high classification accuracy is shown experimentally, and the method's effectiveness is verified by an example from the actual world.

Approach 3

In a case study of credit card fraud detection, the use of Hierarchical Clustering and Artificial Neural Neural Networks yielded promising results, suggesting that, by clustering features, neural inputs might be reduced in the present system. MLP-trained models trained on normalised data have been demonstrated to provide the best results. In this research, we used the unsupervised learning technique. This study's significance rests in its efforts to enhance the accuracy of fraud detection techniques. This study makes use of a dataset created from real-world, non-fictitious commercial dealings processed by a large European company; all personal details have been removed. The average algorithm has a 50% success rate. The key contributions of this research were the identification of an algorithm and the reduction of the cost indicator. In light of this data, they settled on Bayes' minimum risk strategy as yielding the best result (by 23%, to be precise).

Disadvantage:

To better reflect the advantages and disadvantages of fraud detection, we propose a new comparison index in this research.

Two, a cost-conscious Bayesian minimum risk strategy is provided using the proposed cost measure.

In the proposed system, a random forest method is used to categorise the credit card dataset. In both regression and classification, the Random Forest method is often used. It is essentially a collection of several decision trees. When compared to decision trees, random forest performs better since it avoids models from being too specialised for their training data. We randomly choose one subset of the entire feature set to employ as avoid or minimize at each node, allowing us to train each trees separately and then create a decision tree. Random forests can be trained quickly, even for large data sets with many features and examples, since each tree is learned independently. The Random Forest method successfully estimated the generalisation error and showed robustness against over fitting. After that, a hmm was mapped onto the observed and forecasted data.

Advantage: 1. The results generated by the Logit Algorithm are superior than those generated by its rivals.

2.The Logistic algorithm yielded almost ideal outcomes, demonstrating its usefulness.

3.The charts were made once the relevant data had been processed.

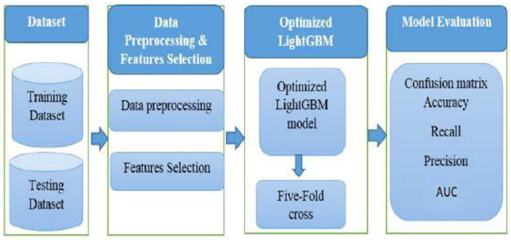


Fig.2: System architecture

MODULES:

Data Collection

The authors of this research utilized the Kaggle dataset for their analysis. The credit card transactions in this data set are extremely skewed, with 492 frauds out of a total of 284,807. It stores the PCA transformation's numerical output variables as input variables. The PCA results in the features V1, V2, V3, V4,......, V28. Because of privacy concerns, we are unable to disclose our proprietary components. Two additional characteristics, time and quantity, have not been PCA converted. The feature "amount" measures the total value of a transaction, whereas the feature "time" measures the amount of time that has passed since the first transaction in the dataset. As the class variable, the response variable is set to 1 if fraud has occurred and 0 otherwise.

Preparing the Data

It's possible that certain key values are missing from the information that was gathered, which might lead to inconsistencies. The effectiveness of the algorithm may be enhanced by preprocessing the data for better results. It is necessary to eliminate outliers and convert variables. We use the map function to solve these problems.

Obtaining Features

identity theft using data collected from Kaggle. After some preliminary data mining, we settled on a logistic regression model to provide reliable results in our reports. As a promising choice for binary classification, logistic regression was used. The project was implemented in Python using the sklearn library; we used Kaggle datasets for Credit card fraud detection; pandas to data frame for class ==0 forno fraud and class==1 for fraud; matplotlib for plotting the fraud and non fraud data; train test split for data extraction (Split arrays or matrices into random train and test subsets); and Logistic Regression machine learning algorithm for fraud detection and print predicting score according to the At last, a Confusion Matrix was constructed by comparing the actual and expected values.

Model Prediction:

Analyzing the data and making a prediction is the goal of predictive modeling. Future occurrences that are now unknown may be predicted using a predictive model. In this procedure, we will develop, validate, and test the model.

Table 1 shows that out of a total of 284,807 transactions, there were 148 instances of fraud. This is an output of predictive modeling, which splits the data set into 70% testing and 30% training. To foresee the result of two days' worth of credit card theft. Unfortunately, 85295 credit card transactions were not fraudulent and needed to be reviewed.

A Retrospective Look at Logical Relationships

In the case of a categorical dependent variable, a regression model known as logistic regression or logit model might be used. Logistic analysis indicates there were 79 fraudulent transactions out of 85,279. To extend a model in which a value of 1 indicates fraud and a value of 0 indicates the absence of fraud, the category variable is class variable may take on the values 0 or 1.

Decision-making Diagram

When constructing a regression model, a decision tree employs a tree structure, producing an end result similar to Fig. 1's tree, where the node at the very top is the root node. A decision node is a collection of two or more branches, where each branch provides a value for the attribute being evaluated, and a leaf node stores the class label. The root node indicates the best predictor in the data. In the case of fraud, a 1 at the leaf node may indicate deception, whereas a 0 indicates normal operation.

Predictive data confusion matrix

When making inferences between pairings, we get a categorical result (-1, 0, or 1) that indicates the frequency with which our predictions match the truth. The sensitivity and precision of the projected data are also determined by this metric. The confusion matrix, sometimes called an error matrix, is a tool in the statistical classification problem used in machine learning.

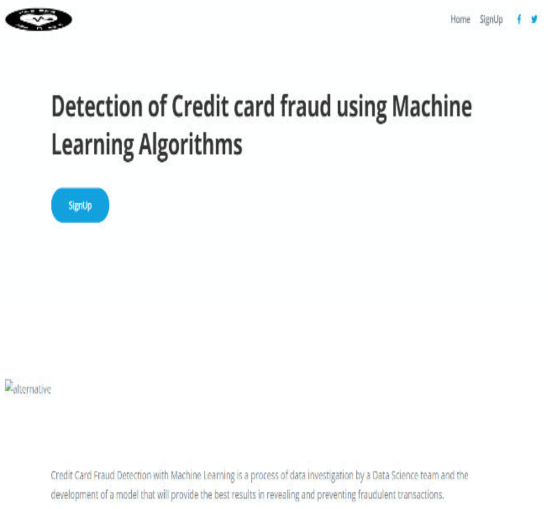
4. IMPLEMENTATION

Naive Bayes: This method of classification is grounded on Bayes' theorem, which states that the existence of a given feature in a class is independent of the presence of another characteristic in a different class. Estimating P(X Y), the probability or probability density of features X given class Y, underpins Naive Bayes classification.

The K-nearest-neighbor (KNN) approach is applicable to both classification and regression issues. We shall employ the KNN for classification and prediction analysis because of its widespread use in the business world. K-Nearest Neighbors (KNN) is a straightforward technique that uses a consensus of its k nearest neighbors to assign a classification to a new instance. According to some distance function, the instance under consideration has the most in common with its K closest neighbors. Distances in Euclidean space, in Manhattan space, in Minkowski space, and in Hamming space are often used.

Estimating true values from estimates of a continuous variable is the purpose of linear regression (s). Linear regression is used to determine a correlation between two sets of data by selecting the best-fitting line. Regression lines are linear equations in which the dependent variable (Y) is represented by the slope (a), the independent variable (X), and the intercept (b). It is through minimizing this total that the coefficients a and b are determined.

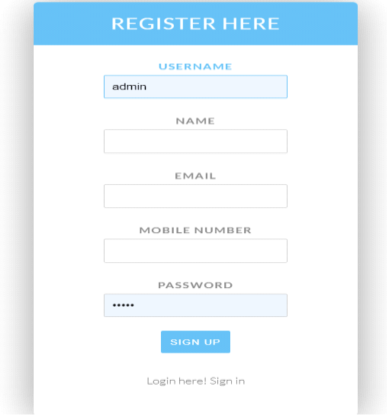
5. EXPERIMENTAL RESULTS



Detection of Credit card fraud using Machine Learning Algorithms

Credit Card Fraud Detection with Machine Learning is a process of data investigation by a Data Science team and the development of a model that will provide the best results in revealing and preventing fraudulent transactions.

Fig.3: Home screen



REGISTER HERE

USERNAME
admin

NAME

EMAIL

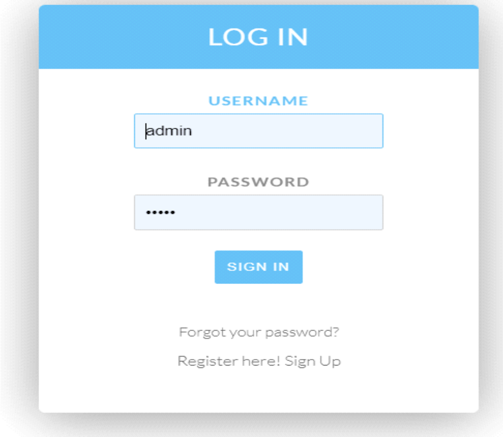
MOBILE NUMBER

PASSWORD
•••••

SIGN UP

Login here! Sign in

Fig.4: User registration



LOG IN

USERNAME
admin

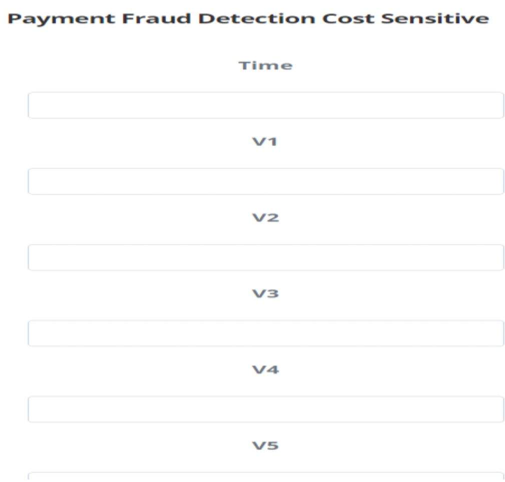
PASSWORD
•••••

SIGN IN

Forgot your password?

Register here! Sign Up

Fig.5: User login



Payment Fraud Detection Cost Sensitive

Time

V1

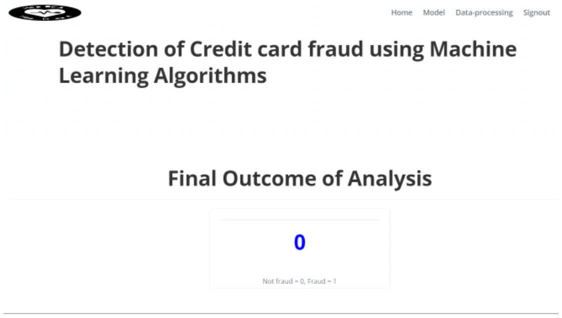
V2

V3

V4

V5

Fig.6: Input screen



Home Model Data-processing Signout

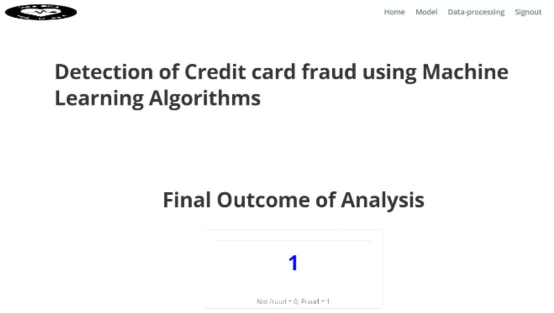
Detection of Credit card fraud using Machine Learning Algorithms

Final Outcome of Analysis

0

Not fraud = 0, Fraud = 1

Fig.7: Prediction result



Home Model Data-processing Signout

Detection of Credit card fraud using Machine Learning Algorithms

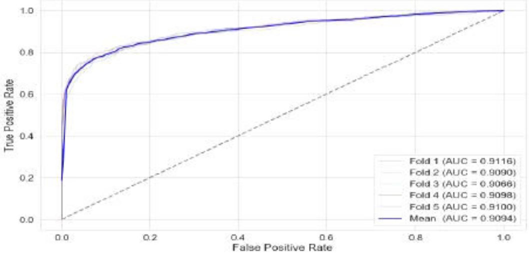
Final Outcome of Analysis

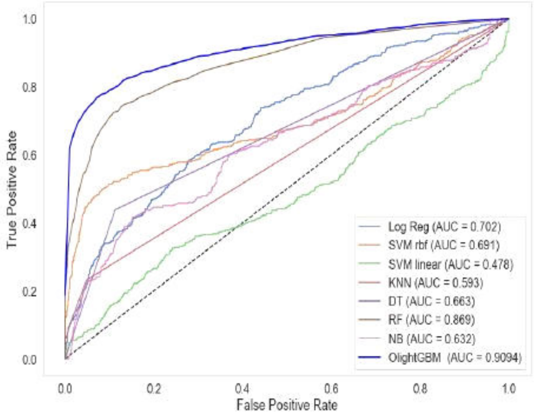
1

Not fraud = 0, Fraud = 1

Fig.8: Prediction result

When compared to other methods, the suggested method yielded the greatest area under the curve (AUC) scores, 90.94% and 92.88% for data sets 1 and 2, respectively. . With an AUC of 47.80% and 70.90% for data sets 1 and 2, respectively, the SVM linear method performed poorly. Accuracy was lowest for the NB method (85% for data set 1 and 95.92% for data set 2), although it was still above chance.





6. CONCLUSION

The capacity to detect and prevent fraudulent transactions has a substantial influence on the growth of credit card use. Since financial institutions are losing significant amounts of money on a regular basis due to credit card fraud, it is crucial to find better ways to detect and prevent this crime. This study proposes a novel way for detecting credit card fraud by using Random Forest. Two real-world datasets were used in our series of experiments. Research findings and province machine learning algorithms were used to evaluate the effectiveness of the proposed approach. These algorithms and methods included: arbitrary forest, regression models, radial vector machine, kernel support vector computer, k-nearest neighbours, decision tree, but also naive bayes. The proposed technique achieved the highest experimental Accuracy, AUC, Precision, and F1-score outcomes. The results show that the proposed algorithm is superior than other approaches. It is demonstrated that the prediction performance of the proposed approach may be enhanced by using an efficient parameter optimization strategy.

7. FUTURE WORK

We want to work on enhancing the model in the future by taking into account the properties of other dimensions beyond frequency, and then apply it to picture data.

REFERENCES

[1] X. Zhang, Y. Han, W. Xu, and Q. Wang, ``HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture,'' Inf. Sci., May 2019. Accessed: Jan. 8, 2019.

[2] N. Carneiro, G. Figueira, and M. Costa, ``A data mining based system for credit-card fraud detection in e-tail,'' Decis. Support Syst., vol. 95, pp. 91101, Mar. 2017.

[3] B. Lebichot, Y.-A. Le Borgne, L. He-Guelton, F. Oblé, and G. Bontempi, ``Deep-learning domain adaptation techniques for credit cards fraud detection,'' in Proc. INNS Big Data Deep Learn. Conference, Genoa, Italy, 2019, pp. 7888.

[4] H. John and S. Naaz, ``Credit card fraud detection using local outlier factor and isolation forest,'' Int. J. Comput. Sci. Eng., vol. 7, no. 4, pp. 10601064, Sep. 2019.

[5] C. Phua, R. Gayler, V. Lee, and K. Smith-Miles, ``On the communal analysis suspicion scoring for identity crime in streaming credit applications,'' Eur. J. Oper. Res., vol. 195, no. 2, pp. 595612, Jun. 2009.

[6] Sudhamathy G: Credit Risk Analysis and Prediction Modelling of Bank Loans Using R, vol. 8, no-5, pp. 1954-1966.

[7] LI Changjian, HU Peng: Credit Risk Assessment for ural Credit Cooperatives based on Improved Neural Network, International Conference on Smart Grid and Electrical Automation vol. 60, no. - 3, pp 227-230, 2017.

[8] Wei Sun, Chen-Guang Yang, Jian-Xun Qi: Credit Risk Assessment in Commercial Banks Based On Support Vector Machines, vol.6, pp 2430-2433, 2006.

[9] Amlan Kundu, Suvasini Panigrahi, Shamik Sural, Senior Member, IEEE,"BLAST-SSAHA Hybridization for Credit Card Fraud Detection", vol. 6, no. 4 pp. 309-315, 2009.

[10] Y. Sahin and E. Duman, "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines, Proceedings of International Multi Conference of Engineers and Computer Scientists, vol. I, 2011.

[11] Sitaram patel, Sunita Gond , "Supervised Machine (SVM) Learning for Credit Card Fraud Detection, International of engineering trends and technology, vol. 8, no. -3, pp. 137- 140, 2014.

[12] Snehal Patil, Harshada Somavanshi, Jyoti Gaikwad, Amruta Deshmane, Rinku Badgujar," Credit Card Fraud Detection Using Decision Tree Induction Algorithm, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.4, April- 2015, pg. 92-95

[13] Dahee Choi and Kyungho Lee, "Machine Learning based Approach to Financial Fraud Detection Process in Mobile Payment System", vol. 5, no. - 4, December 2017, pp. 12-24.