Exploiting The Machine-Learning Techniques for Intensifying the Credit Card Fraud Detection

Dr.M.Chinna rao[1], Ms. Sravya[2], Mr. Harsha vardhan[3],Ms. Chetana[4], Ms. Indu priya[5]
[1,2,3,4,5]*Department of Computer Science and Engineering/Lingayas Institute of Management and Technology/JNTUK/INDIA*

Abstract- In this paper, mainly focused on credit card fraud detection for in real world. Initially collect the credit card datasets for trained dataset. Billions of dollars of loss are caused every year by fraudulent credit card transactions[1]. The design of efficient fraud detection algorithms is the key for reducing these losses, and more and more algorithms rely on advanced machine learning techniques to assist fraud investigators. The design of fraud detection algorithms is however particularly challenging due to the non-stationary distribution of the data[2.3], the highly unbalanced classes distributions and the availability of few transactions labeled by fraud investigators. Then will provide the user credit card queries for testing data set. After classification process of random forest algorithm using to the already analyzing data set and user provide current dataset[4.5]. Finally optimizing the accuracy of the result data. Then will apply the processing of some of the attributes provided can find affected fraud detection in viewing the graphical model visualization. The performance of the techniques is evaluated based on accuracy, sensitivity, and specificity, precision[6]. The results indicate about the optimal accuracy for Decision tree are 98.6% respectively.

Keywords: credit card fraud detection, random forest algorithm, fraud detection, visualization, Decision tree.

I. INTRODUCTION

Now a days the usage of credit cards has dramatically increased. Credit card fraud detection is a very popular but also a difficult problem to solve. As credit card becomes the most popular mode of payment for both online as well as regular purchases cases of fraud associated with it are also rising.began replacing manual imprinters[7.8]. POS terminals allowed a business to electronically capture and send credit card information, greatly speeding up the transaction its approval process, and contributing to the growth in credit card usage[9.10].Creditcard fraud happens when someone a fraudster or a thief uses your stolen credit card or the information from that card to make unauthorized purchases in your name[10]. When an individual uses someone else's card information illegally or for uninformed personal spending, it is classified as credit card fraud[11]. Credit card fraud happens when someone a fraudster or a thief uses your stolen credit card or the information from that card to make unauthorized purchases in your name. When an individual uses someone else's card information illegally or for uninformed personal spending, it is classified as credit card fraud.

Mainly credit card frauds are classified into four types.

-Pick pocketing or Physical theft.
-Skimming card information.
-Phishing and other scams.
-Carding or Cyberattacks.

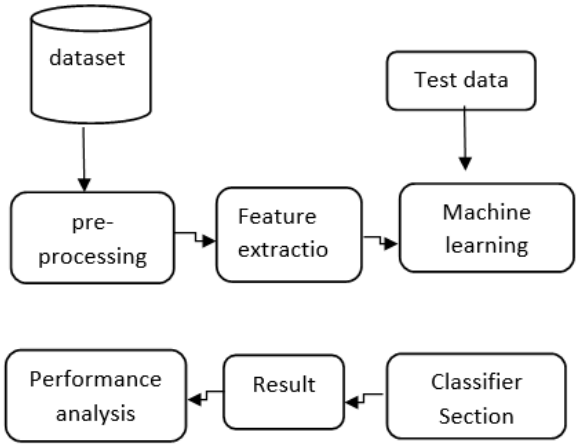
Bank transactions as well as credit card frauds increased. One of the most target frauds are credit card fraud, the fraud can occur any type of credit products, such products are retail, home loan and personal loan. During the last few decades as technology has changed, dramatically the face of fraud also changed. To detect credit card fraud data mining techniques, Predictive modeling and Logistic Regression are used. In prediction model to predict the continuous valued functions. Credit card of CSV files will be analyzed to predict the outcome[12].

Credit cards truly entered the mainstream with widespread acceptance in the 1960s when Bank of America (eventually *Visa*) introduced the general-purpose credit card. IBM developed magnetic stripe technology in 1969incorporating it into credit cards[13].

In this paper, we model the sequence of operations in credit card transaction processing using a Decision tree, Random Forest, support vector machines and

show how it can be used for the detection of frauds. Algorithms are initially trained. with the normal behavior of a cardholder. If an incoming credit card transaction is not accepted by the trained with sufficiently high probability, it is fraudulent. At the same time, we try to ensure that genuine transactions. We present detailed experimental results to show the effectiveness of our approach and compare it with other techniques available in the literature[14].

II. SYSTEM ARCHITECTURE



Above System architecture represents how the frauds are detected is described. Data set is a collection of reviews collected from existed transaction records. These datasets are stored in csv files. In this proposed system the dataset was downloaded from the Kaggle which contains nearly 2,84,000 transactions. Data preprocessing step converts the raw data into a well-developed format which is useful for analysis part by removing the unwanted data. Feature extraction methods are used to remove unnecessary, irrelevant, and redundant attributes from a dataset that do not increase the accuracy. Then train the dataset by selecting a machine learning models like decision tree. After applying classifiers, by comparing the results based on the accuracy of the model's algorithm will be applied. The efficient credit card fraud detection system will be implemented which classifies whether the transaction is genuine or fraud.

If you are using Word, use either the Microsoft Equation Editor or the MathType add-on (http://www.mathtype.com) for equations in your paper (Insert | Object | Create New | Microsoft Equation or MathType Equation). ―Float over text‖ should not be selected[17.18].

III.MATERIALS AND METHODS

Existing system:

In existing system methods such as Cluster Analysis, Support vector machine (SVM), Bayesian network, Logistic Regression, Naïve Bayer's, Hidden Markov model (HMM) e.t.c are used to find out the credit card fraud transactions which are unstable on large data sets. The methods used in the existing system which are based on unsupervised learning[15].

Proposed system:

The proposed techniques are used for detecting the frauds in credit card system. The comparison made for different machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest, to determine which algorithm suits best and can be adapted by credit card merchants for identifying fraud transactions[16]. In this proposed model system is developed by using the supervised machine learning techniques as well as unsupervised techniques for an efficient performance.

Advantages:

- More accurate result
- Able to detect different fraudulent behavior.
- Cost and Time efficient.

Software requirements:

- Operating system: Windows8/10
- Coding: Python 3.6
- API's: Numpy, pandas, scikit.

Hardware requirements:

- Processor: corei3 2.1GHZ
- RAM: 4GB(minimum)
- Hard disk drive: 10GB(free)

In credit card fraud detection, supervised machine learning techniques like random forest, support vector machines, decision trees are applied.

1.DECISION TREES ALGORITHM:

Decision Tree Analysis is a general, predictive modelling tool that has applications spanning a number of different areas. In general, decision trees are constructed via an algorithmic approach that

identifies ways to split a data set based on different conditions. It is one of the most widely used and practical methods for supervised learning. Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A decision tree is a tree-like graph with nodes representing the place where we pick an attribute and ask a question; edges represent the answers the to the question; and the leaves represent the actual output or class label. They are used in non-linear decision making with simple linear decision surface. Decision trees classify the examples by sorting them down the tree from the root to some leaf node, with the leaf node providing the classification to the example. Each node in the tree acts as a test case for some attribute, and each edge descending from that node corresponds to one of the possible answers to the test case. This process is recursive in nature and is repeated for every subtree rooted at the new nodes. Root node describes the best predictor in the data, and decision node is a combination of two or more branches, each branch represents a value for the attribute which is tested, leaf node holds class label. The leaf node may be 1 means fraud and 0 otherwise. A decision tree is a tree structure. The decision tree can be linearized into decision rules, easy to understand.

2. Support Vector Machine algorithm:
Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

3. Random Forest algorithm:
Random forest is a supervised machine learning algorithm based on ensemble learning. Ensemble learning is an algorithm where the predictions are derived by assembling or bagging different models or similar model multiple times. The random forest algorithm works in a similar way and uses multiple algorithm i.e multiple decision trees, resulting in a forest of trees, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks.
Advantages of using random forest:
The random forest algorithm is not biased and depends on multiple trees where each tree is trained separately based on the data, therefore biasedness is reduced overall. It is a very stable algorithm. Even if a new data point is introduced in the dataset, it does not affect the overall algorithm rather affect the only a single tree. It works well when one has both categorical and numerical features.

IV.RESULTS

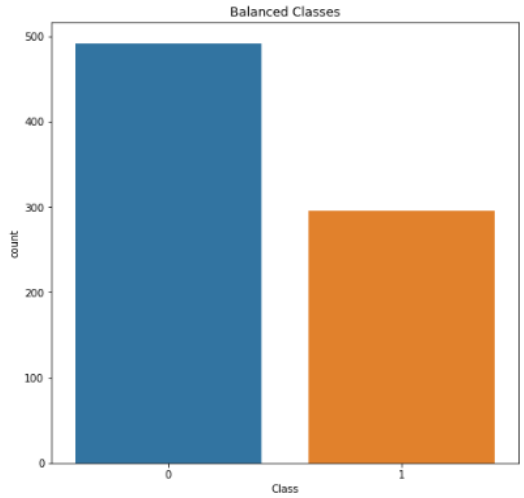


Fig.2Data Exploration

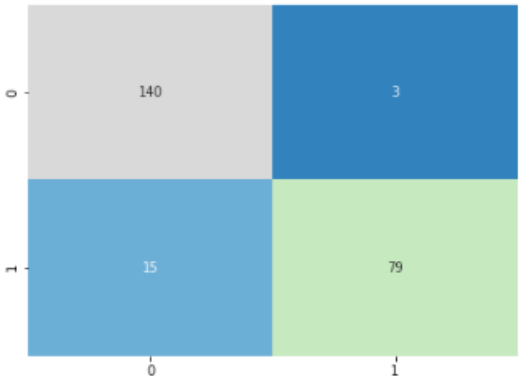


Fig.3 SVM confusion matrix

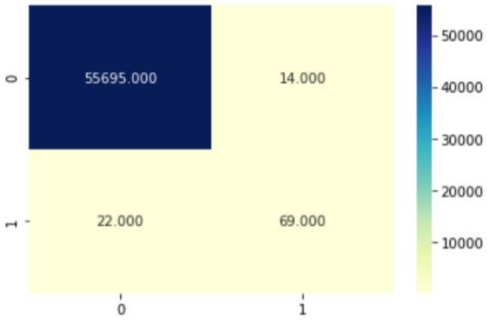


Fig.4Random Forest confusion matrix

V.CONCLUSION

In this analysis of credit card fraud detection using machine learning has been discussed in detail. The design of one such system has also been done. Credit card fraud detection system is in high demand as online banking increases day by day, cybercrime related to online payment frauds also increases. These systems are crucial for the creation of a better experience for the customers.

It has been found that the performance of these systems depends highly on the type of model that is used for classification. Several algorithms such as SVM, Random Forest, Logistic regression.

This fraud detection system model helps the people to predict the credit card transaction which is genuine or fraud. In this proposed system, random forest outperforms the remaining algorithms. The proposed paper evaluates that the random forest, Decision tree and support vector machine algorithm will perform better with a larger number of training data, but speed during testing and application will suffer. Application of more pre-processing techniques would also help. The SVM algorithm still suffers from the imbalanced dataset problem and requires more pre-processing to give better results at the results shown by SVM is great but it could have been better if more pre-processing have been done on the data.so, in proposed work we balanced the imbalanced data with up-sampling technique during pre-processing. We review the existing works on credit card fraud prediction in three different perspectives: datasets, methods, and metrics. Firstly, we present the details about the availability of public datasets and what kinds of details are available in each dataset for predicting credit card fraud. Secondly, we compare the various predictive modeling methods that have been used in the literature for predicting, and then quantitatively compare their performances in terms of accuracy.

REFERENCES

[1] Y. Sahin and E. Duman, “Detecting Credit Card Fraud by Decision Trees and Support Vector Machines,” International Multiconference of Engineers and computer scientists, vol. 1, pp. 442–447, 2011.

[2] R.Dorronsoro,F.Ginel,S.Carmen,andC.S.Cruz,“ NeuralFraudDetection in Credit Card Operations,” IEEE Transactions on Neural Networks, vol. 8, no. 4, pp. 827–834, 1997.

[3] K. Ramakalyani and D. Umadevi, “Fraud Detection of Credit Card Payment System by Genetic Algorithm,” International Journal of Scientific & Engineering Research, vol. 3, no. 7, pp. 1–6, 2012.

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

[5] Saloni, Minakhi Rout, “Analysis and Comparisonof Credit Card Fraud Detection Using Machine Learning” in Advances in Electronics, Communication and Computing, 2021, pp. 33-40.

[6] A.Srivastava,A.Kundu,S.Sural,andS.Member,“C reditCardFraudDetection Using Hidden Markov Model,” IEEE Transactions on Dependable and Secure Computing, vol. 5, no. 1, pp. 37–48, 2008.

[7] J. S. Mishra, S. Panda, and A. K. Mishra, “A Novel Approach for Credit Card Fraud Detection Targeting the Indian Market,” International Journal of Computer Science, vol. 10, no. 3, pp. 172–179, 2013.

[8] R. J. Bolton and D. J. Hand, “Statistical fraud detection: A review,” Statistical science, pp. 235–249, 2002.

[9] R. Patidar and L. Sharma, “Credit Card Fraud Detection Using Neural Network,” International Journal of Soft Computing and Engineering, no. May, pp. 13–14, 2011.

[10] D. J. Weston, D. J. Hand, N. M. Adams, and C. Whitrow, “Plastic card fraud detection using peer group analysis,” vol. 2, pp. 45–62, 2008.

[11] E. Duman and M. H. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search," Expert Systems with Applications, vol. 38, no. 10, pp. 13057– 13063, 2011.
[12] K. Ramakalyani and D. Umadevi, "Fraud Detection of Credit Card Payment System by Genetic Algorithm," International Journal of Scientific & Engineering Research, vol. 3, no. 7, pp. 1–6, 2012.
[13] P. J. Bentley, J. Kim, G.-h. Jung, and J.-u. Choi, "Fuzzy Darwinian Detection of Credit Card Fraud," pp. 1–4,2007.
[14] A.Srivastava, A.Kundu, S.Sural, and S.Member, "Credit Card Fraud Detection Using Hidden Markov Model," IEEE Transactions on Dependable and Secure Computing, vol. 5, no. 1, pp. 37–48, 2008.
[15] S. Esakkiraj and S. Chidambaram, "Predictive Approach for Fraud Detection Using Hidden Markov Model," International Journal of Engineering Research & Technology, vol. 2, no. 1, pp. 1–7, 2013.
[16] J. S. Mishra, S. Panda, and A. K. Mishra, "A Novel Approach for Credit Card Fraud Detection Targeting the Indian Market," International Journal of Computer Science, vol. 10, no. 3, pp. 172–179, 2013.
[17] A. Brabazon, J. Cahill, P. Keenan, and D. Walsh, "Identifying Online Credit Card Fraud using Artificial Immune Systems," IEEE Congress on Evolutionary Computation, pp. 1 – 7, 2010
[18] N. Wong, P. Ray, G. Stephens, and L. Lewis, "Artificial immune systems for the detection of credit card fraud: an architecture, prototype and preliminary results," Information systems, vol. 22, pp. 53–76, 2012.