International Journal of Research in Engineering, Science and Management
Volume-2, Issue-6, June-2019
www.ijresm.com | ISSN (Online): 2581-5792

Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection

N. S. DeepaShree[1], S. Vijaya Lakshmi[2], Tejashwini Alagundagi[3], R. Bhumika[4], Sunitha Myageri[5]
[1]*Assistant Professor, Dept. of Computer Science and Engg., Global Academy of Technology, Bengaluru, India*
[2,3,4,5]*UG Student, Dept. of Computer Science and Engg., Global Academy of Technology, Bengaluru, India*

***Abstract*: Detecting frauds in credit card transactions is perhaps one of the best test beds for computational intelligence algorithms. In fact, this problem involves a number of relevant challenges, namely: concept drift, class imbalance, and verification latency. However, the vast majority of learning algorithms that have been proposed for fraud detection rely on assumptions that hardly hold in a real-world fraud-detection system. This lack of realism concerns two main aspects: The way and timing with which supervised information is provided and the measures used to assess fraud-detection performance. The goal of data analytics is to delineate hidden patterns and use them to support informed decisions in a variety of situations. Credit card fraud is escalating significantly with the advancement of modernized technology and became an easy target for frauds. Credit card fraud has highly imbalanced publicly available datasets. In this project, we apply many supervised machine learning algorithms to detect credit card fraudulent transactions using a real-world dataset. Furthermore, we employ these algorithms to implement a super classifier using ensemble learning methods. We identify the most important variables that may lead to higher accuracy in credit card fraudulent transaction detection. Additionally, we compare and discuss the performance of various supervised machine learning algorithms that exist in literature against the super classifier that we implemented in this work.**

***Keywords*: supervised machine learning algorithms, fraudulent transaction, real world dataset, fraud detection.**

1. Introduction

Machine learning is an approach which has the ability to understand the data already trained by the users to machine to enact like human and take its own decision by experience and give best prediction results. Machine learning is a booming technology which helps people to reduce their works and the machine itself act like human and give best result for every user given data it may be an images, audio, videos, text, huge amount of data, document, etc. Machine learning experiences the set of task to improve the performance of system to give best result and predict the things.

The figure 1 shows how the machine itself trained with dataset to get experience by performing some tasks. The dataset may be anything it may unstructured data or structured data.

In this project proposed a supervised machine learning algorithms for credit card fraudulent transaction detection. Three machine learning algorithms are proposed namely: random forest, Naïve Bayes and k Nearest Neighbour. Proposed a comparative analysis of three algorithms, Prediction and detection of the fraudulent in the transaction.

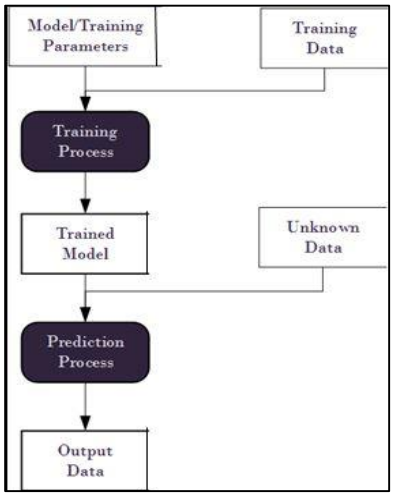

Fig. 1. Machine learning training process

2. Literature survey

A. Study of hidden markov in credit card fraudulent detection

V. Bhusari [1] proposed study of hidden markov in credit card fraudulent detection. In this project classified an user profiles into three types namely; lower profile, middle profile and higher profile. Based on these user profiles it detects fraudulent. It keeps record of spending profile of the card holder by both way, either offline or online. Thus analysis of purchased commodities of cardholder will be a useful tool in fraud detection system and it is assuring way to check fraudulent transaction, although fraud detection system does not keep records of number of purchased goods and categories.

B. Credit card fraud identification using artificial neural network

In first phase login credential and credit card details are checked. If validation check is passing then transaction is passed to second phase otherwise transaction will be rejected. In second phase neural network training based learning is performed. Transaction is fraudulent or genuine is decided by second phase [2]. If transaction is fraudulent then rejected and if genuine then allowed by second phase.

C. Credit card fraud detection using machine learning models and collating machine learning models

Logistic Regression is a supervised classification method that returns the probability of binary dependent variable that is predicted from the independent variable of dataset that is logistic regression predict the probability of an outcome which has two values either zero or one, yes or no and false or true. Logistic regression has similarities to linear regression but as in linear regression a straight line is obtained, logistic regression [3].

D. Cost sensitive modelling of credit card fraud using neural network strategy

Proposed two methods namely; neural network with multi-layer structure and meta cost procedure. The Artificial Neural Network was presented by Warren S. McCulloch and Walter Pitts for classification and prediction problems. Multilayer Perceptron (MLP) [4] is a supervised technique with one input layer, one or more hidden layers and one output layer in which every layer involves some neurons, which is a feed forward neural net (there is no feedback loop).

3. Methodology

A system architecture is the conceptual model that defines the structure, behavior and more views of the system. An architecture description is the formal description and representation of the system, organized in a way that supports reasoning about the structures and behavior of the system. The system architecture can consist of system components and the subsystems developed that will work together to implement the overall system. There have been efforts to formalized languages to describe system architecture.

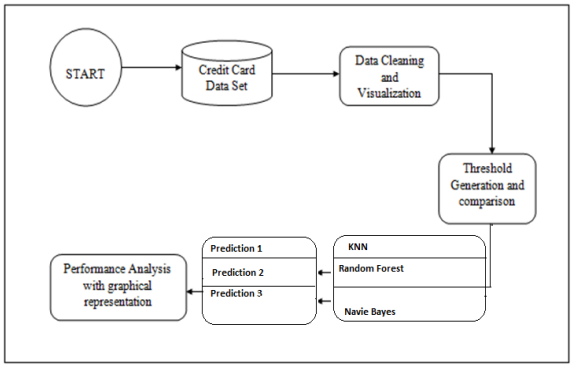

Fig. 2. Analysis and prediction of fraudsters using data set

The Figure 2 depicts, the system architecture is going to start by taking the credit card data set, that means it contain the transactions made by the credit cards.

- The credit card dataset, is processed for data cleaning and data visualization.
- Data Cleaning-This is the process of detecting and correcting the corrupt records from the database, tables or records and by replacing and modifying or deleting the data.
- It uses three algorithms KNN, Random Forest and Naive Bayes, the output of these two is provided as prediction 1 and 2. The analysis takes place and represented in a graphical manner.

4. Experimental analysis

This section describes the screens of the "Supervised machine Learning Algorithms for Credit Card Fraudulent Transaction Detection". The output result snapshots are shown below for each module.

The Fig. 3 displays the front page of credit card fraud system which displays all the algorithms and the accuracy.

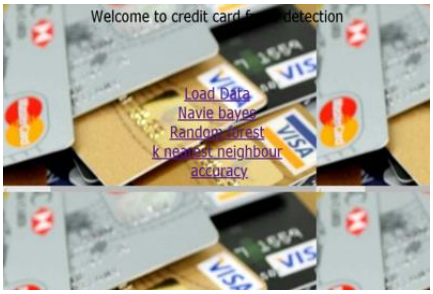

Fig. 3. Main page

- Fetching dataset by clicking show dataset as shown in figure 4.

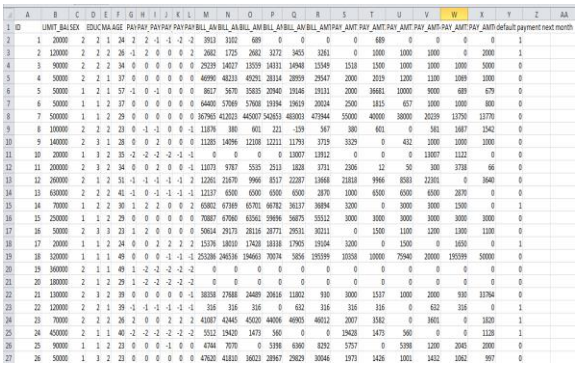

Fig. 4. Displays dataset

- The below figure 5 will display the Likely hood table of Naive Bayes algorithm of the selected attributes.
- Based on the likely hood table and the accuracy of Naive Bayes algorithm is being generated along with graph.

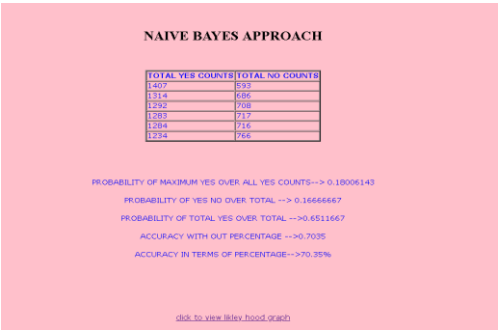

Fig. 5. Likely hood table

- Figure 6 shows naïve bayes yes count no count graph for PAY_0 to PAY_6

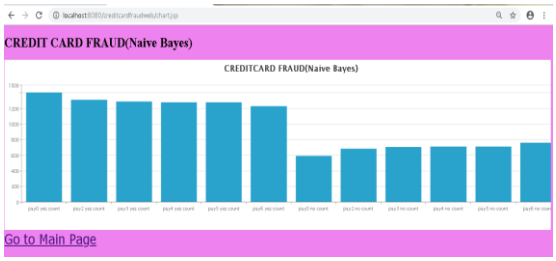

Fig. 6. Naive bayes graph

- The Figure 7 depicts that, the fraudulents and non fraudulents predicted by applying the random forest algorithm.

960:1040

USER	STATUS
USER 1	non fraudlent
USER 2	non fraudlent
USER 3	fraudlent
USER 4	fraudlent
USER 5	fraudlent
USER 6	fraudlent
USER 7	fraudlent
USER 8	non fraudlent
USER 9	non fraudlent
USER 10	non fraudlent
USER 11	non fraudlent
USER 12	fraudlent

Fig. 7. Random forest prediction

- The Figure 8 depicts that, the fraudulent and non -fraudulent predicted by applying the random forest algorithm in the form of graph.

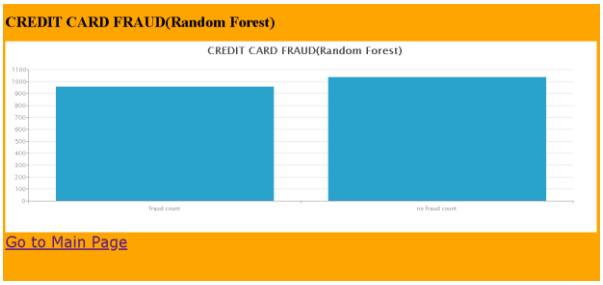

Fig. 8. Random forest prediction graph

- The Figure 9 depicts that entering values from pay_0 to pay_6 for prediction from KNN.

Fig. 9. KNN new tuple entry

- The figure 10 depicts that, the count of fraudsters and non-fraudsters by using KNN algorithm.

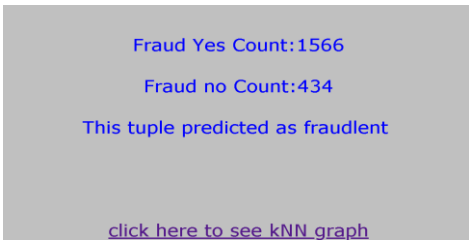

Fig. 10. KNN prediction

- The Figure 11 depicts that, the fraudulent and non-fraudulent predicted by applying the kNN algorithm in the form of graph.

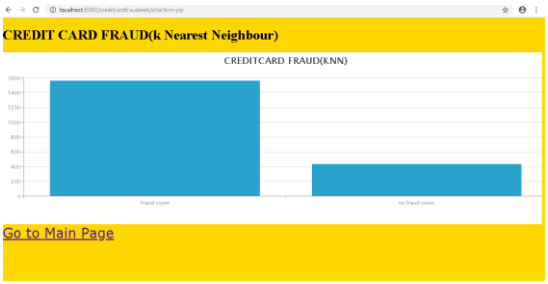

Fig. 11. KNN algorithm graph

- The above figure indicates the comparison of the three algorithms with the graph.

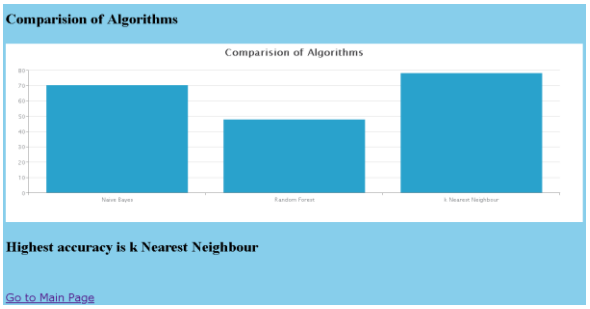

Fig. 12. Comparison of all algorithms

5. Conclusion and future work

In this project, we applied three supervised machine learning algorithms to detect credit card fraudulent transactions using a real-world dataset. This is due to the fact that clustering relies entirely on the similarities and differences of features of the dataset. As the k nearest neighbour showed the highest accuracy than the Random Forest algorithm and Naïve Bayes algorithm we can say that or conclude that k nearest neighbour is the best among the three algorithms.

Bank software or application can be developed for the detection of credit card fraudulent transaction. Many more machine learning supervised algorithm can be added to get best accuracy for the detection of fraud.

References

[1] Bhusari, S. Patil, "Study of Hidden Markov Model in Credit Card Fraudulent Detection",2016. IEEE/IAFE Computat. Intell. Financial Eng., Mar. pp. 220–226.
[2] Chandrahas MishraA, Dharmendra Lal GuptaB, Raghuraj SinghC "Credit Card Fraud Identification using Artificial Neural Network", 2015.vol. 24, no. 8, pp. 791–800.
[3] Navanshu Khare and Saad Yunus Sait, "Credit Card Fraud Detection using Machine Learning Models and Collating Machine Learning Model", vol. 28, no. 2, pp. 246–258, Feb. 2016.
[4] Fahimeh Ghobadi and Mohsen Rohani, "Cost Sensitive Modeling of Credit Card Fraud Using Neural Network Strategy",2016., vol. 24, no. 4, pp. 620–634, Apr. 2013.
[5] B. Baesens, V. Van Vlasselaer, and W. Verbeke, Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection. Hoboken, NJ, USA: Wiley, 2015.
[6] Gupta, Shalini, and R. Johari, "A New Framework for Credit Card Transactions Involving Mutual Authentication between Cardholder and Merchant," International Conference on Communication Systems and Network Technologies IEEE, 2011:22-26.
[7] Y. Gmbh and K. G. Co, "Global online payment methods: Full year 2016," Tech. Rep., 3 2016.
[8] Bolton, Richard J., and J. H. David, "Unsupervised Profiling Methods for Fraud Detection," Proc Credit Scoring and Credit Control VII (2001):5–7.
[9] Seyedhossein, Leila, and M. R. Hashemi. "Mining information from credit card time series for timelier fraud detection," International Symposium on Telecommunications IEEE, 2011:619-624.
[10] Srivastava, A., Kundu, A., Sural, S., and Majumdar, A, "Credit card fraud detection using hidden markov model," IEEE Transactions on Dependable and Secure Computing, 5(1), 37-48, 2008.