JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 10, ISSUE 04, 2023

CREDIT CARD FRAUD DETECTION USING RANDOM FOREST AND CART ALGORITHM

Mrs. M.SYAMALA SAISREE[1], K.SRICHANDANA[2], K.AISHWARYA[3], K. HARIKA[4]

[1]Assistant Professor, Department of Information Technology Malla Reddy Engineering College for Women (UGC-Autonomous) Maisammaguda,Hyderabad-500100

[2,3,4]UG Students, Department of Information Technology Malla Reddy Engineering College for Women (UGC-Autonomous) Maisammaguda,Hyderabad-500100

ABSTRACT: Billions of dollars of loss are caused every year by fraudulent credit card transactions. The design of efficient fraud detection algorithms is 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, the highly unbalanced class distributions, and the availability of few transactions labeled by fraud investigators. At the same time, public data are scarcely available for confidentiality issues, leaving unanswered many questions about what the best strategy is. In this thesis, we aim to provide some answers by focusing on crucial issues such as) why and how under-sampling is useful in the presence of class imbalance (i.e. frauds are a small percentage of the transactions), ii) how to deal with unbalanced and evolving data streams (non-stationarity due to fraud evolution and change of spending behavior), iii) how to assess performances in a way which is relevant for detection and iv) how to use feedbacks provided by investigators on the fraud alerts generated. Finally, we design and assess a prototype of a Fraud Detection System able to meet real-world working conditions and that can integrate investigators' feedback to generate accurate alerts.

Keywords: CNN, fraud detection, secure data, feedback

INTRODUCTION

Online shopping growing day to day. Credit cards are used for purchasing goods and services with the help of virtual card cards physical card whereas virtual cards are for online transactions and physical cards are for offline transactions. In a physical card-based purchase, the cardholder presents his card physically to a merchant for making a payment. To carry out fraudulent transactions in this purchase, an attacker must steal the credit card. If the cardholder does not realize the loss of the card, it can lead to a substantial financial loss to the credit card company. In online payment mode, attackers need only a little information for doing fraudulent transactions (secure code, card number, expiration date, etc.). This purchase method will mainly be done through the Internet or telephone. To commit fraud in these types of purchases, a fraudster simply needs to know the card details. Most of the time, the genuine cardholder is not aware that someone else has seen or stolen his card information. The only way to detect this kind of fraud is to analyze the spending patterns on every card and to figure out any

inconsistency concerning the "usual" spending patterns. Fraud detection based on the analysis of existing purchase data of cardholders is a promising way to reduce the rate of successful credit card fraud. Since humans tend to exhibit specific behaviorist profiles, every cardholder can be represented by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent, etc. Deviation from such patterns is a potential threat to the system.

LITERATURE SURVEY

1)The Use of Predictive Analytics Technology to Detect Credit Card Fraud in Canada (Kosemani Temitayo Hafiz, Dr. Shaun Aghili, Dr. Pavol Svirsky)

This research paper focuses on the creation of a scorecard from relevant evaluation criteria, features, and capabilities of predictive analytics vendor solutions currently being used to detect credit card fraud. The scorecard provides a side-by-side comparison of five credit card predictive analytics vendor solutions adopted in Canada. The ensuing research findings outlined a list of credit card fraud PAT vendor solution challenges, risks, and limitations.

2) BLAST-SSAHA Hybridization for Credit Card Fraud Detection (Amlan Kundu, Suvasini Panigrahi, Shamik Sural, Senior Member, IEEE, and Arun K. Majumdar)

In this paper, we propose to use a two-stage sequence alignment in which a profile Analyser (PA) first determines the similarity of an incoming sequence of transactions on a given credit card with the genuine cardholder's past spending sequences. The unusual transactions traced by the profile analyzer are next passed on to a deviation analyzer (DA) for possible alignment with past fraudulent behavior. The final decision about the nature of a transaction is taken based on the observations by these two analyzers. To achieve online response time for both PA and DA, we suggest a new approach for combining two sequence alignment algorithms BLAST and SSAHA.

3)Research on Credit Card Fraud Detection Model Based on Distance Sum

(Wen-Fang YU, Na Wang)

Along with increasing credit cards and growing trade volume in China, credit card fraud rises sharply. How to enhance the detection and prevention of credit card fraud becomes the focus of risk control of banks. This paper proposes a credit card fraud detection model using outlier detection based on distance sum according to the infrequency and unconventionality of fraud in credit card transaction data, applying outlier mining into credit card fraud detection. Experiments show that this model is feasible and accurate in detecting credit card fraud.

4)Fraudulent Detection in Credit Card Systems Using SVM & Decision Tree (Vijayshree B. Nipane, Poonam S. Kalinga, Dipali Vidhate, Kunal War, Bhagyashree P. Deshpande)

With growing advancements in the electronic commerce field, fraud is

spreading all over the world, causing major financial losses. In the current scenario, the Major cause of financial losses is credit card fraud; it not only affects trades person but also individual clients. Decision trees, Genetic algorithms, Meta-learning strategy, neural networks, and HMM are the presented methods used to detect credit card fraud. In contemplating systems for fraudulent detection, the artificial intelligence concept of Support Vector Machine (SVM) & decision tree is being used to solve the problem. Thus, by the implementation of this hybrid approach, financial losses can be reduced to a greater extent.

5)Supervised Machine (SVM) Learning for Credit Card Fraud Detection (Sitaram Patel, Sunita Gond)

In this thesis, we are proposing the SVM (Support Vector Machine) based method with multiple kernel involvement which also includes several fields of user profile instead of only spending profile. The simulation result shows improvement in TP (true positive), and TN (true negative) rates, & also decreases the FP (false positive) & FN (false negative) rates.

EXISTING SYSTEM

This was on k-means Algorithm implementation, Only the two features with the most variance were used to train the model. The model was set to have 2 clusters, 0 being non-fraud and 1 being a fraud. We also experimented with different values for the hyperparameters, but they all produced similar results. Changing the dimensionality of the data (reducing it to more dimensions than 2) also made little difference in the final values.

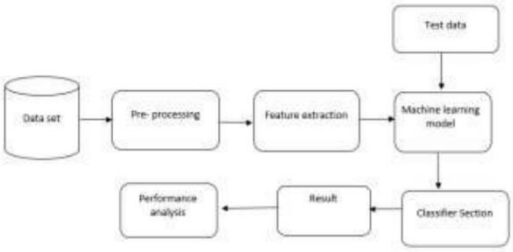

Fig.1. Architecture model

PROPOSED SYSTEM

Our goal is to implement a machine learning model to classify, to the highest possible degree of accuracy, credit card fraud from a dataset gathered from Kaggle. After initial data exploration, we knew we would implement a logistic regression model for best accuracy reports. Logistic regression, as it was a good candidate for binary classification. Python learns library was used to implement the project, We used Kaggle datasets for Credit card fraud detection, using pandas to a data frame for class ==0 for no 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 used Logistic Regression machine learning algorithm for fraud detection and print predicting score according to the algorithm. Finally, the Confusion matrix was plotted on true and predicted.

Module Description:

- **Upload credit card dataset**: we collected the credit card fraud data from the Kaggle website.
- **Generate train and test model**: we must pre-process the collected data by cleaning null values, unwanted rows, and unwanted columns. After that, we have to split the data into two parts training part with 80% and the testing part with 20%.
- **Run the Random Forest algorithm**: we must train the training data for the RF algorithm and test with test data to get accuracy.
- **Detect Credit card Fraud from test data**: by using random forest we can detect the fraud signatures.
- **Clean and fraud graph**: we can display the clean and fraudulent transactions with a graph.

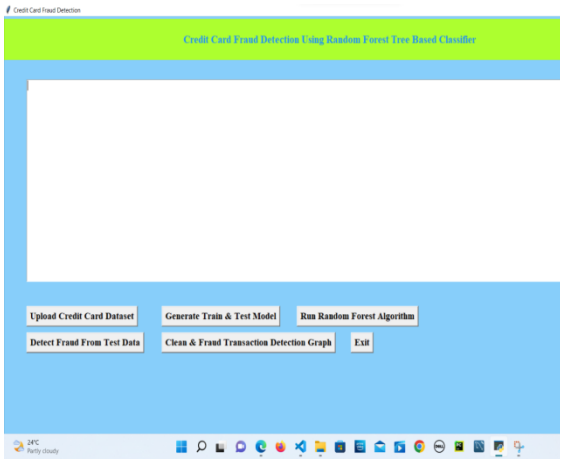

Fig 1 Home page.

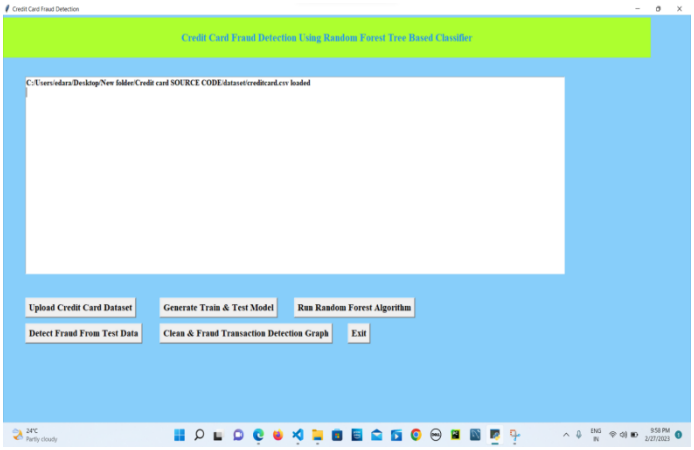

Fig.2 upload credit card data set

Upload credit card data set: we collected the credit card fraud data from the Kaggle website.

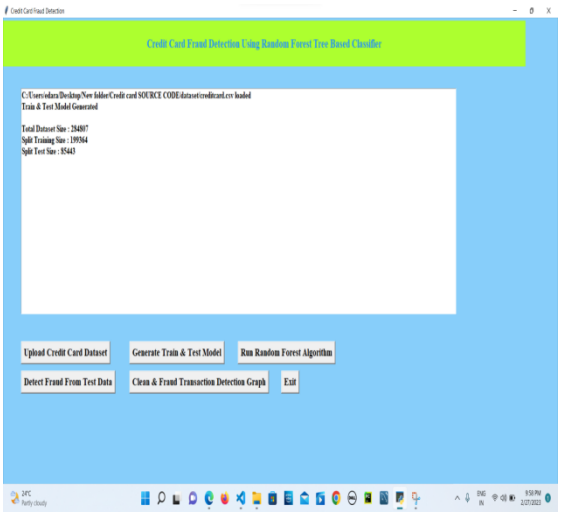

Fig 3 Generate Train & Test Model

Generate train & test model: we must pre-process the collected data by cleaning null values, unwanted rows, and unwanted columns. After that, we have to split the data into two parts training part with 80% and the testing part with 20%.

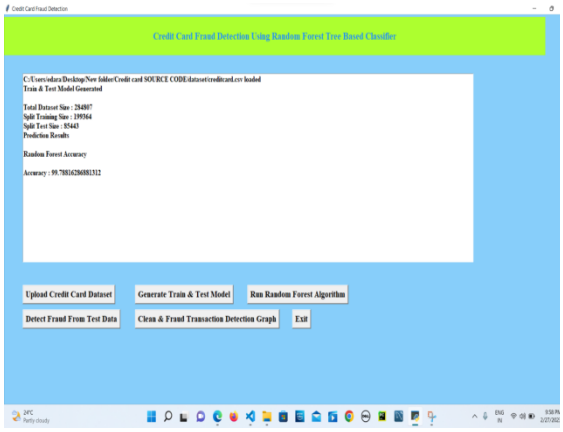

Fig 3 Run Random Forest Algorithm.

Run Random Forest Algorithm: we must train the training data for the RF algorithm and test with test data to get accuracy.

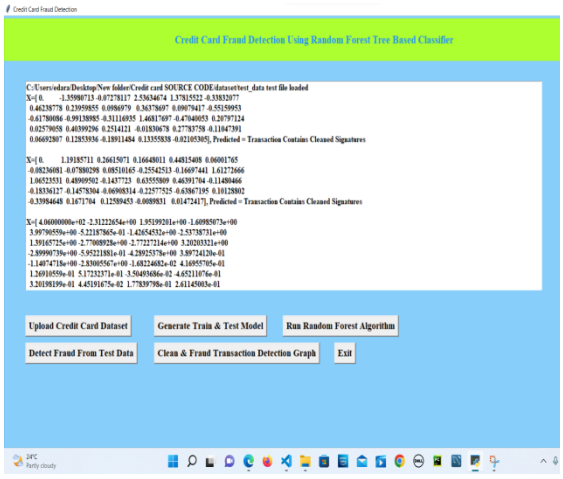

Fig 4 Detect Fraud from Test Data

Detect Fraud from Test Data: By using random forest we can detect the fraud signatures.

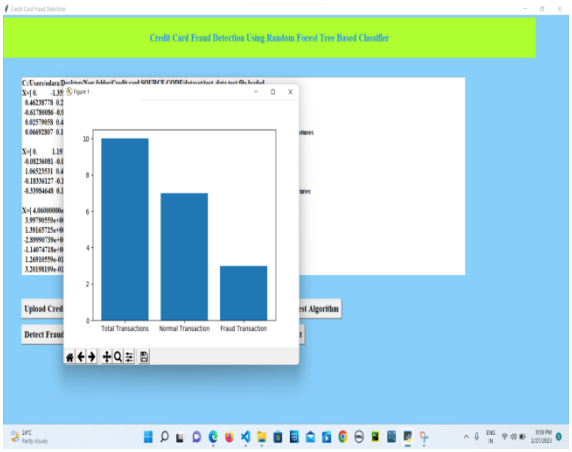

Fig 5 clean and fraud graph.

CONCLUSION

The Random Forest algorithm will perform better with a larger number of training data, but speed during testing and application will suffer. The application of more pre-processing techniques would also help. The SVM algorithm still suffers from the imbalanced dataset problem and requires more preprocessing to give better results at the results shown by SVM are great, but it could have been better if more preprocessing have been done on the data.

FUTURE ENHANCEMENT

Our future work will focus on solving these problems. The algorithm of the random forest itself should be improved. For example, the voting mechanism assumes that each of the base classifiers has equal weight, but some of them may be more important than others. Therefore, we also try to make some improvements to this algorithm

REFERANCES

- ✓ Salazar, Addison, et al. "Automatic credit card fraud detection based on non-linear signal processing." Security Technology (ICCST), 2012 IEEE International Carnahan Conference on. IEEE, 2012.
- ✓ Delamaire, Linda, H. A. H. Abdou, and John Pointon. "Credit card fraud and detection techniques: a review." Banks and Bank systems 4.2 (2009): 57-68.
- ✓ Quinlan, J. Ross. "Induction of decision trees." Machine learning 1.1 (1986): 81-106.
- ✓ Quinlan, J. R. (1987). "Simplifying decision trees". International Journal of Man-Machine Studies. **27** (3): 221. doi:10.1016/S0020-7373(87)80053-6.
- ✓ K. Karimi and H.J. Hamilton (2011), "Generation and Interpretation of Temporal Decision Rules", International Journal of Computer Information Systems and Industrial Management Applications, Volume 3.
- ✓ Aggarwal, Charu C. "Outlier analysis." Data mining. Springer International Publishing, 2015.
- ✓ Salazar, Addisson, Gonzalo Safont, and Luis Vergara. "Surrogate techniques for testing fraud detection algorithms in credit card operations." Security Technology (ICCST), 2014 International Carnahan Conference on. IEEE, 2014.

Ogwueleka, Francisca Nonyelum. "Data mining application in credit card fraud detection system." Journal of Engineering Science and Technology 6.3 (2011): 311-322