iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** III **Month of publication:** March 2019

DOI: http://doi.org/10.22214/ijraset.2019.3215

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

Credit Card Fraud Detection using Random Forest Algorithm

Gokula Krishnan. V[1], Dhinesh Raj. S[2], Lokesh. S[3], Sudharshan. S[4]
[1]*Associate Professor, Department of Computer Science and Engineering Panimalar Institute of Technology, Chennai, Tamil Nadu, India.*
[2, 3, 4]*UG Scholars, Department of Computer Science and Engineering, Panimalar Institute of Technology, Chennai, Tamil Nadu, India.*

Abstract: ***Technological developments have changed the way we live. Banks have introduced the concept of credit cards. Due to the advancement in the electronic commerce technology, the use of credit cards has increased and it has become a popular mode of payment for both online and offline purchases.***
In spite of this enormous popularity the cards are not free of risk.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. Our project mainly focussed on credit card fraud detection in real world.
Initially we will collect the credit card datasets for trained dataset. Then we will provide the user credit card queries for testing data set. After classification process of dataset random forest algorithm is used for analysing data set and current dataset provided by the user. After final optimization the results indicates about the optimal accuracy for Random Forest Algorithm is 98.6% of the accuracy of the result data.
Keywords: FDS (Fraud Detection System),Credit Card, Random Forest Algorithm

I. INTRODUCTION

Credit cards are widely used due to the popularization of ecommerce and the development of mobile intelligent devices. Credit card has made an online transaction easier and more convenient. Fraud detection is a process of monitoring the transaction behaviour of a cardholder in order to detect whether an incoming transaction is done by the cardholder or others. Credit card fraud detection is a relevant problem that draws the attention onal intelligence communities, where a large number of automatic solutions have been proposed. In a real-world FDS, the massive stream of payment requests is quickly scanned by automatic tools that determine which transactions to authorize.

Classifiers are typically employed to analyze all the authorized transactions and alert the most suspicious ones. Alerts are then inspected by professional investigators that contact the cardholders to determine the true nature (either genuine or fraudulent) of each alerted transaction. By doing this, investigators f machine-learning and computatio provide a feedback to the system in the form of labelled transactions, which can be used to train or update the classifier, in order to preserve (or eventually improve) the fraud-detection performance over time.

The vast majority of transactions cannot be verified by investigators for obvious time and cost constraints. These transactions remain unlabeled until customers discover and report frauds. Another important difference between what is typically done in the literature and the real-world operating conditions of Fraud-Detection System (FDS) concerns the measures used to assess the fraud-detection performance.

We use random forest to train the normal and fraud behaviour features. Random forest is a classification algorithm based on the votes of all base classifiers.

II. RELATED WORK

A comprehensive understanding of fraud detection technologies can be helpful for us to solve the problem of credit card fraud. The work in [5] evaluates two advanced data mining approaches, support vector machines and random forests, together with the well-known logistic regression, as part of an attempt to better detect (and thus control and prosecute) credit card fraud. The study is based on real-life data of transactions from an international credit card operation. The work in [7] is a study which employs transaction aggregation strategy to detect credit card fraud.

III. LITERATURE SURVEY

SL.NO	TITLE	AUTHOR	YEAR	CONCEPT	ADVANTAGES	DISADVANTAGES
1	Real Time Credit Card Fraud Detection using Computational Intelligence	Jon T. S. Quah and M. Sriganesh	2008	Real-time fraud detection is a major concern. A new and innovative approach in understanding spending patterns to decipher potential fraud cases is used. It makes use of Self Organization Map to decipher, filter and analyze customer behaviour for detection of fraud.	Transactions are categorized and only potential fraudulent cases are reviewed manually which helps in reduction of overall cost and time.	Pattern matching technique used in the existing system can be tracked easily.
2	Learning from Imbalanced Data	Haibo He, and Edwardo A. Garcia	2009	A comprehensive review of the development of research in learning from imbalanced data is being used. The focus is to provide a critical review of the nature of the problem, the state-of-the-art technologies, and the current assessment metrics used to evaluate learning performance under the imbalanced learning scenario.	Decision making process becomes easier and overall performance is improved.	Analysis from raw data to support decision-making processes has become difficult. Performance of learning algorithms is affected.
3	Association rules applied to credit card fraud detection	D. Sanchez, M.A. Vila, L. Cerda, J.M. Serrano	2009	Association rules are used in order to extract knowledge so that normal behaviour patterns may be obtained in unlawful transactions from transactional credit card databases in order to detect and prevent fraud.	The problem of minimum support and confidence is overcome due to which fraud detection becomes easier.	Decision making process is difficult in credit card fraud detection.
4	A hybrid model for plastic card fraud detection systems	M. Krivko	2010	Hybrid model for plastic card fraud detection system is being used. The ability of the hybrid model to identify fraudulent activity on the real debit card transaction data is demonstrated.	Hybrid model is capable of identifying fraudulent activity in a timely manner resulting in substantial monetary savings.	Detecting potentially fraudulent activity on a debit card account is a major concern.
5	Data mining for credit card fraud: A comparative study	Siddhartha Bhattacharyya, SanjeevJhab, KurianTharakunnel, J. Christopher Westland	2011	It evaluates two advanced data mining approaches, and support vector machines and random forests, together with the well-known logistic regression, as part of an attempt to better detect (and thus control and prosecute) credit card fraud.	Logistic regression shows a noteworthy performance rate and low fraudulent rate.	Lack of availability of data search for the predictive model.
6	Incremental Learning of Concept Drift in Non stationary Environments	Ryan Elwell, and RobiPolikar	2011	Classifiers-based approach is being used for incremental learning of concept drift, characterized by non stationary environments (NSEs). Learn++ .NSE, learns from consecutive batches of data without making any assumptions on the nature or rate of drift.	This is applicable for non stationary environment. NSE can track the changing environments very closely, regardless of the type of concept drift.	The existing model is only applicable in stationary environment and it requires access to the previous data.
7	Employing Transaction Aggregation Strategy To Detect Credit Card Fraud	SanjeevJhaa, Montserrat Guillenb, J. Christopher Westland	2012	Transaction aggregation strategy is being employed to detect credit card fraud. Aggregated transactions are to capture consumer buying behaviour prior to each transaction and used these aggregations for model estimation to identify fraud-lent transactions.	Analysis of customer buying behaviour is possible and it helps to achieve high of performance rate.	The existing model cannot identify customer buying behaviour prior of the transaction done by the customer.
8	Just-In-Time Classifiers for Recurrent Concepts	CesareAlippi, GiacomoBoracchi, and Manuel Roveri	2013	A novel generation of JIT classifiers is presented which deals with recurrent concept drift by means of a practical formalization of the concept representation and the definition of a set of operators working on such representations.	JIT classifier improves its accu-racy over time by exploiting additional supervised information coming from the field.	Low accuracy and the existing model cannot deal with the recurrent concept drift.

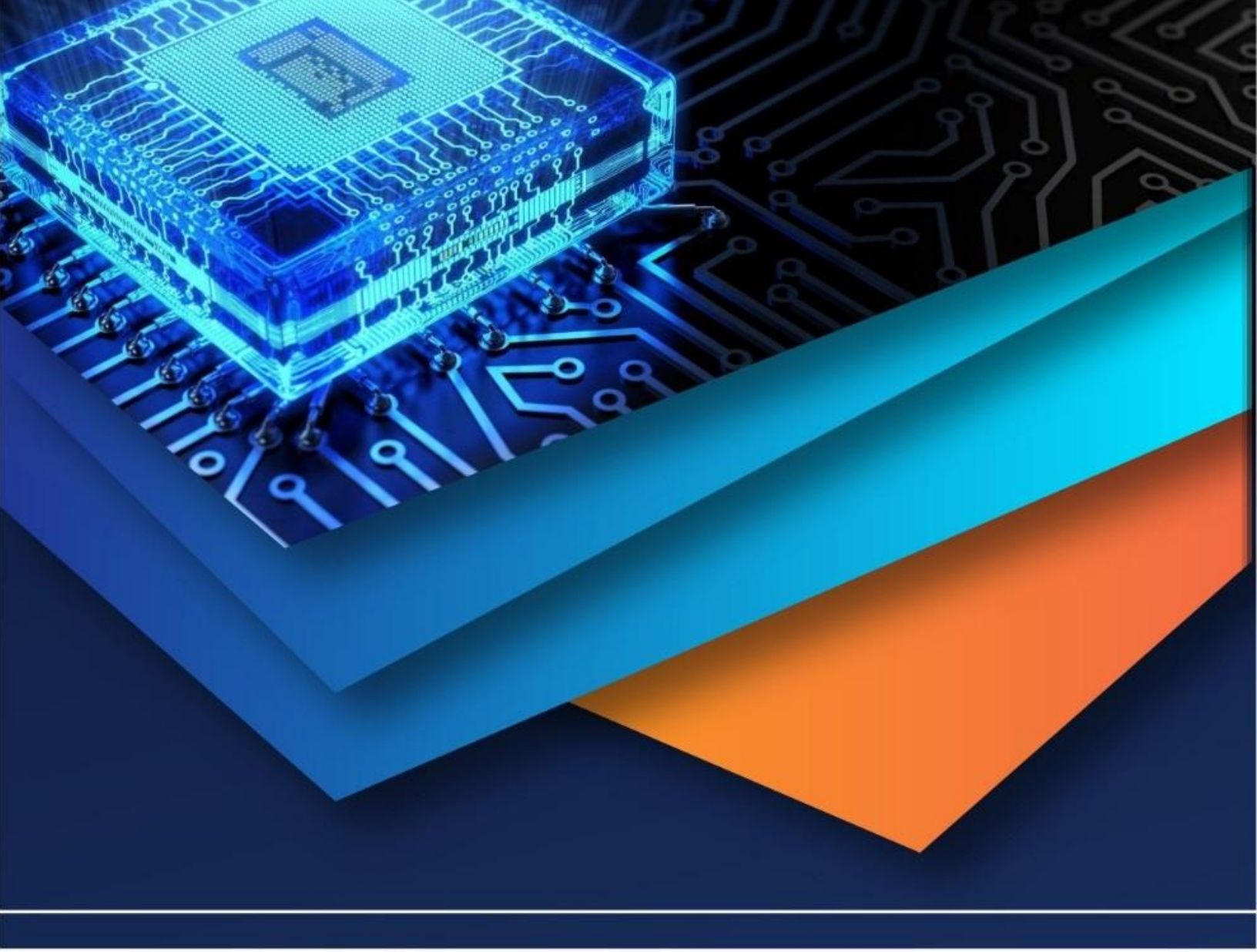
9	Bank Sealer: An Online Banking Fraud Analysis and Decision Support System	Michele Carminati, Roberto Caron, Federico Maggi, IleniaEpifani, Stefano Zanero	2014	A semi-supervised online banking fraud analysis and decision support approach is proposed. During the training phase, it builds a profile for each customer based on past transactions.	The under training due to the lack of historical data for building of well-trained profiles, and the users that change their habits is mitigated.	The existing system does not contain historical data of the customers for building of well-trained profiles.
10	Detecting Credit Card Fraud using Periodic Features	Alejandro Correa Bahnsen, DjamilaAouada, AleksandarStojanovic and Bjorn Ottersten	2015	To create a new set of features based on analyzing the periodic behaviour of the time of a transaction using the von Mises distribution technique is being proposed. The state-of-the-art credit card fraud detection models, compares and evaluates how the different sets of features have an impact on the results.	The proposed method uses periodic data for evaluation. By including the proposed periodic features into the methods, the results show an average increase in savings of 13%.	Usage of raw transactional features, such as time, amount, and place of the transaction and the spending behaviour of the customer is not taken into account.
11	Credit Card Fraud Detection and Concept-Drift Adaptation with Delayed Supervised Information	Andrea Dal Pozzolo, GiacomoBoracchi, Olivier Caelen, CesareAlippi and GianlucaBontempi	2015	A realistic fraud-detection setting is addressed and it is shown that the investigator's feedbacks and delayed labels have to be handled separately and two FDSs on the basis of an ensemble and a sliding-window approach are being designed.	Alert precision, which is the primary concern of investigators, can be substantially improved by the proposed approach.	In the existing system the Fraud Detection System (FDS) faces difficulty in detecting frauds due to concept drift and class imbalance.
12	Calibrating Probability With Under Sampling for Unbalanced Classification	Andrea Dal Pozzolo, Olivier Caelent, Reid A. Johnson, GianlucaBontempi	2015	How under sampling affects the posterior probability of a machine learning model is studied both analytically and experimentally. Bayes Minimum Risk theory is being used to find the correct classification threshold and show how to adjust it after under sampling. An experiment on several real-world unbalanced datasets is done to validate the results.	The proposed result gives high classification accuracy and probability calibration.	The under sampling technique significantly impacts the classification accuracy and probability calibration.
13	Detecting credit card fraud by Modified Fisher Discriminant Analysis	Nader Mahmoudi, EkremDuman	2015	A linear discriminant, called Fisher Discriminant Function is used for investigation for the first time in credit card fraud detection problem. A Modified Fisher Discriminant Function is proposed which makes the traditional function more sensitive to the important instances.	Profit that can be obtained from a fraud/legitimate classifier is maximized and the experimental results confirm that Modified Fisher Discriminant could eventuate more profit.	Supervised learning methods uses complex algorithm which over fit the data set they are built on and this classification does not concentrate on more important instances.
14	Hierarchical Change-Detection Tests	CesareAlippi, GiacomoBoracchi, and Manuel Roveri	2016	Hierarchical change-detection tests (HCDTs) are being used, as an effective online algorithms for detecting changes in data streams. Experiments has shown that, when the process generating the data stream is unknown, as it is mostly the case in the real world.	HCDTs achieve a far more advantageous trade off between false-positive rate and detection delay and has low false positive rate.	The existing system as high false positive rate and high delay in fraud detection. Identification of changes in data stream is difficult.
15	Credit Card Fraud Detection: A Realistic Modelling and a Novel Learning Strategy	Andrea Dal Pozzolo, GiacomoBoracchi, Olivier Caelen, CesareAlippi, and GianlucaBontempi,	2018	Fraud-detection problem is being formalized that realistically describes the operating conditions of FDSs that everyday analyzes massive streams of credit card transactions. Random Forest algorithm is being used which effectively addresses the issues such as class imbalance, concept drift, and verification latency.	The problems such as concept drift class imbalance latency verification have been overcome.	The main problem faced in the existing system model for fraud detection are concept drift, class imbalance and latency verification.

IV. CONCLUSION

This paper has analysed the performance of Random Forest Algorithm in credit card fraud detection system and the final optimization results indicates the optimal accuracy for Random Forest Algorithm is 98.6%. Although random forestobtains good results on given data set, there are still someproblems such as imbalanced data. Our future work will focuson solving these problems.

REFERENCES

[1] J. T. Quah and M. Sriganesh, "Real-time credit card fraud detection using Computational intelligence," Expert Syst. Appl., vol. 35, no. 4, pp. 1721–1732, 2008.
[2] H. He and E. A. Garcia, "Learning from imbalanced data," IEEE Trans. Knowl. Data Eng., vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[3] D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, "Association rules applied to credit card fraud detection," Expert Syst. Appl., vol. 36, no. 2, pp. 3630–3640, 2009.
[4] M. Krivko, "A hybrid model for plastic card fraud detection systems," Expert Syst. Appl., vol. 37, no. 8, pp. 6070–6076, 2010.
[5] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," Decision Support Syst., vol. 50, no. 3, pp. 602–613, 2011.
[6] R. Elwell and R. Polikar, "Incremental learning of concept drift in non stationary environments," Trans. Neural Netw., vol. 22, no. 10, pp. 1517–1531, 2011.
[7] S. Jha, M. Guillen, and J. C. Westland, "Employing transaction aggregation strategy to detect credit card fraud," Expert Syst. Appl., vol. 39, no. 16, pp. 12650–12657, 2012.
[8] C. Alippi, G. Boracchi, and M. Roveri, "Just-in-time classifiers for recurrent concepts," IEEE Trans. Neural Netw. Learn. Syst., vol. 24, no. 4, pp. 620–634, Apr. 2013.
[9] M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero, BankSealer: A Decision Support System for Online Banking Fraud Analysis and Investigation, Berlin, Germany: Springer, 2014, pp. 380–394.
[10] C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Detecting credit card fraud using periodic features," in Proc. 14th Int. Conf. Mach. Learn. Appl., Dec. 2015, pp. 208–213.
[11] Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection and concept-drift adaptation with delayed supervised information," in Proc. Int. Joint Conf. Neural Netw., 2015, pp. 1–8.
[12] Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Cali-brating probability with undersampling for unbalanced classification," in Proc. IEEE Symp. Ser. Computat. Intell., Dec. 2015, pp. 159–166.
[13] N. Mahmoudi and E. Duman, "Detecting credit card fraud by modified fisher discriminant analysis," Expert Syst. Appl., vol. 42, no. 5,pp. 2510–2516, 2015.
[14] Alippi, G. Boracchi, and M. Roveri, "Hierarchical change-detection tests," IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 2, pp. 246–258, Feb. 2016.
[15] Andrea Dal Pozzolo, GiacomoBoracchi, Olivier Caelen, CesareAlippi, Fellow, IEEE, and GianlucaBontempi, Senior Member, IEEE, "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy", IEEE Transactions On Neural Networks And Learning Systems, vol 10, pp 216-23, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY