Available online @ https://jjem.jnnce.ac.in
https: www.doi.org/10.37314/JJEM.SP0217
Indexed in International Scientific Indiexing (ISI)
Impact factor: 1.395 for 2021-22
Published on: 08 December 2024

Enhancing security in payment platform to prevent fraud

Ruchitha H.P [1]*, Dr Raghavendra S.P [2]

[1] Student, Assistant Professor, [2] Department of Computer Application
[3] JNN College of Engineering, Shimogga
ruchithaparmesh1122@gmail.com, raghusp@jnnce.ac.in

Abstract

To enhance one's consumer experiences and reduce loss of funds, monetary institutions must aggressively detect transaction risks. In this paper, we compare various machine learning methods for accurately and efficiently predicting the validity of financial transactions. The methods utilized in this work included MLP Repressor, Random Forest Classifier, Complement NB, MLP Classifier, Gaussian NB, Bernoulli NB, LGBM Classifier, Ada Boost Classifier, K Neighbors Classifier, Logistic Regression, Bagging Classifier, Decision Tree Classifier, and Deep Learning. The random forest classifier performed best with unbalanced datasets the accuracy is 97% the precession is 88% the recall rate is 89% and the score for f1 is 95% however using bagging classification performed best on a balanced dataset the accuracy precession recall and f1-score are all 95%.The dataset was collected from Kaggle repository. It consists of 6000 rows and 10 columns. Dataset name is online fraud

Keywords: *Fraud detection, Decision tree, Random Forest, Machine Learning, Anaconda prompt*

1. Introduction

For several years fraudulent transactions with detect-tors played a combined function illegal transactions happen far more often than ever before particularly in today's internet era and they are the leading cause of loss of money transaction crime lost the economy almost a billion dollars in 2019. 30 billion dollars in 2020 and well over than 32 billion dollars in 2021 the rate of around the world fraud on transactions is predicted to increase year after year reaching 34 bill-lion by 2022 as a result finance company and banks might need a digital fraud identification tool to identify and screen money funds scam sur-veil lance system assess payment information to look for strange trends and follow incoming transactions ml refers to an automated intelligence ai technological advances that lets machinery to develop and grow upon prior expertise while been directly built ma-chines learning refers to a innovation of software programs can obtain instruction and then apply them to learn for itself the key objective are to get computer grow freely without human involvement and support and change their activities accordingly de-spite the evidence that training under supervision has been quite effective in detecting forged funds the development new transactional security analysis al-girths will never stop a small improvement to the classifier will save a corporation a significant amount of money the paper will look at research findings or practical instances in which predictive models have been successfully used to detect and prevent fraud related to systems by combining these findings the study hopes to add to the continuing discussion about improving safety measures and protecting monetary transactions in the digital age finally the study emphasizes the value of proactive and advanced ap-poaches to fraud detection in financial payment sys-terms financial institutions may reduce risks safe-guard.

2. Literature Survey

Mohammed et al., [1] in his work gives Card fraud datasets include challenges such as high-class imbalance, labeled and unlabeled samples, and the requirement to analyze a huge number for negotiations, prompting innovative strategies. Real-time suspicion of fraud employs predictive machine learning techniques that consistof choice trees, Bayesian Sorting, Least Squares The process of regression, linguistic regression, and SVM.

Xuan et al., [2] in his work Random Forest lessons deeper than random-tree-based vs cart-based draws near have been employed to teach the behavioral components of normal and anomalous transactions.

Rubio et al., [3]in his work The piece covers the phenomenon concerning flawed information resulting in leads to an excess of unidentified findings also suggests tricks for confronting them.

S. Geetha et al., [4] in his work svm was used to evaluate transfers as valid or a fraud the svm examined the cardholders prior transaction habits when a new transaction occurs the system deviates from its usual behavior by labeling it as illegal the greatest finding fraud score by using svm was 91

A. Thennakoon et al., [5] in his work we use generally accessible emulated settlement activity information to apply several autonomous data mining techniques for catching fraud we want to show how supervised machine-learning methods can probably be utilized to categorize evidence of high class imbalance and high accuracy we show that analytical methods can be utilized to help differentiate between deceptive and nonfraudulent dealings.

Pumsirirat et al., [6] in his work The authors stated the main intent during the dissertation is exploring predictive modeling methods the strategies used involve the ada boost which algorithm commonly and the random forest algorithm the outcomes are reliability precision recall et the f1 rating provide the basis for algorithms the matrix of misinformation serves as the foundation for constructing the roc curve when comparing the random forest and ada boost algorithms the scheme exhibiting the highest degree of accuracy pinpointing recall and the f1 value is considered the most successful for fraud detection.

Aditya Oza et al., [7] in his work algorithms derived from machine learning have been used quite efficiently for recognizing payment-related criminal activity probabilistic approaches allow for finding novel instances of unlawful activity in the present work we use a data set with labels of electronic payment transactions to introduce different computational methods that utilize logistic reconstruction and aid vector machines to the challenge of discovering cash payments fraud their presented techniques efficiently observe suspicious transactions in excellent reliability along with few false positives

T.Singh et al., [8] in his work The authors offer an svm-based manner that can recognize crystalline malicious software the investigation also addresses the prospect of lopsided assortments ie malicious file samples in lieu of asymptomatic paperwork and how to adequately identify those with pinpointing and pinpointing accuracy.

I.Sadgali et al., [9] in his work The authors examined the performance concerning monetary fraud prediction strategies centered around unsupervised data collection artificial intelligence will and sophisticated acquiring knowledge and pitched hybrid designs that would address real-time obstacles.

R. Rambola et al., [10] in his work Statistical procedures might be employed to spot a scam the statistical breakdown with the historical data is analyzed to feed unexpected actions of items illegal utilizing linear discriminant analysis and logistic regression

3. Proposed Methodology

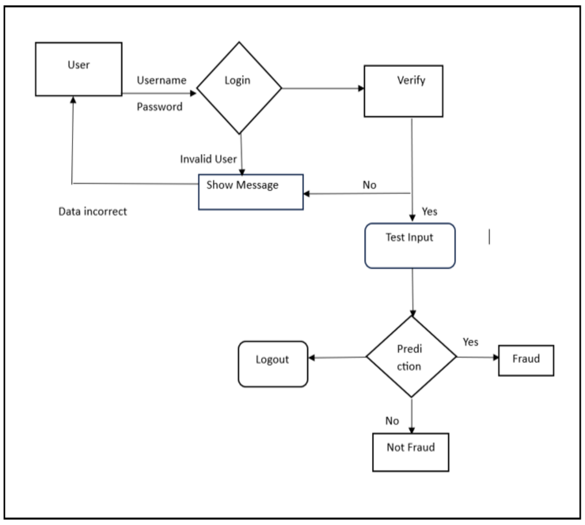

Fig. 1: Block diagram for proposed

methodology

When a user wishes to use a financial payment service they must first enter their username and password on the login page after submission the system uses these credentials to validate the users identity if the login information is correct the user is granted access and a welcome message appears if the data is incorrect the user receives an error and must re-enter their credentials the system uses complex algorithms to examine test inputs and user behavior these algorithms evaluate the login attempt to patterns of previous fraudulent activity if the system detects something unexpected such as several failed login attempts or login it uses a prediction model to determine the possibility of fraud users are encouraged to log out after completing their transactions

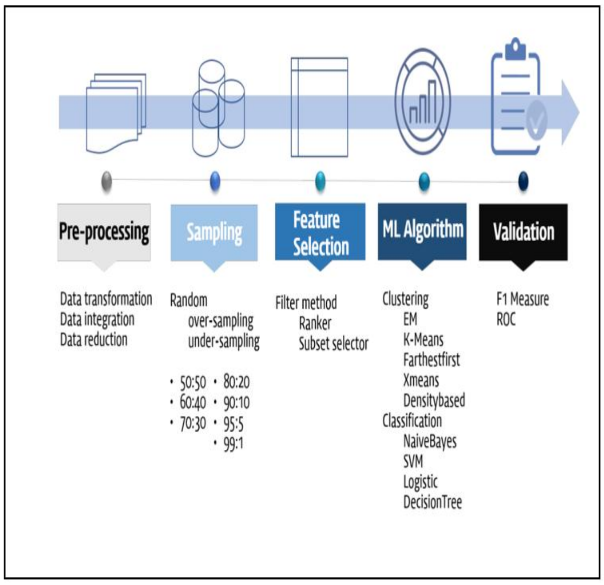

Fig 2: System architecture

3.1 Decision Tree

In decision tree any transaction has been classified as fraud or not fraud and it includes numerous details such as the transaction value a decision tree is built by asking a series of questions about the characteristics that define the transactions resulting in splits that best distinguish fake transactions from non-fraudulent ones the tree assesses all alternative attributes and thresholds and chooses the one that best categorizes activities as mainly fraudulent or mostly non fraudulent this approach determines the best split using a metric such as information gain.

Mathematical equation:

$$IG(S,A) = -\sum_{i=1}^{c} pi\ log2(pi) - \sum_{v€Values(A)} \frac{|Sv|}{|S|} \ldots.(1)$$

Where**:**

- S denotes the initial list of transactions.
- A symbolizes the attribute being considered while splitting.
- Pi represents the probability of class i (fraud or non-fraud) in the original set S.

- Values(A) follow the special contents of identify A.
- Sv is a subsection of S for the aspect A has value v.
- |Sv| shows the number of instances in the subset.
- |S| gives the number of instances in the original set S.
- c reflects the number of classes (two for binary classification: fraud or not fraud).

3.2 Random Forest

Random forests were developed to enhanced the correctness and its durability of identifying scams associated with transaction processing by generating multiple choice trees that are trained on distinct randomly selected pieces of data combining all tree projections to reach a final choice reducing over-fitting and making higher quality forecasts this renders randomized forest analysis a promising method for detecting forged funds and strengthening the privacy of banking and payment services.

$$y\hat{} = mode(\{Ti(x) | i = 1, 2, \ldots, N\})^{\ldots.(2)}$$

where:

- The final projected our lessons over

intake x

- Is symbolized by y whereas Ti(x) is the projected length achieved by the i-th most likely choice in the forest
- N is the whole number of choices throughout the chaotic forest
- Mode denotes the greatest proportion of the forecasts among every selection trees

Tools Used

- Django
- Bootstrap
- Numpy
- Pandas

4. Results and Discussion

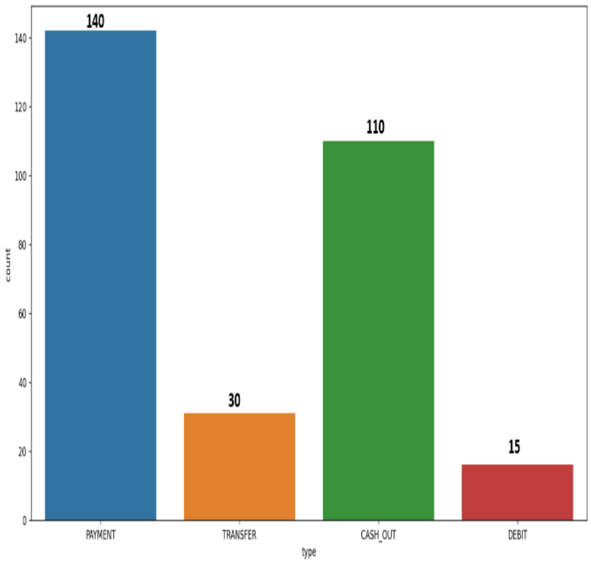

Fig3: The bar graph of financial transaction

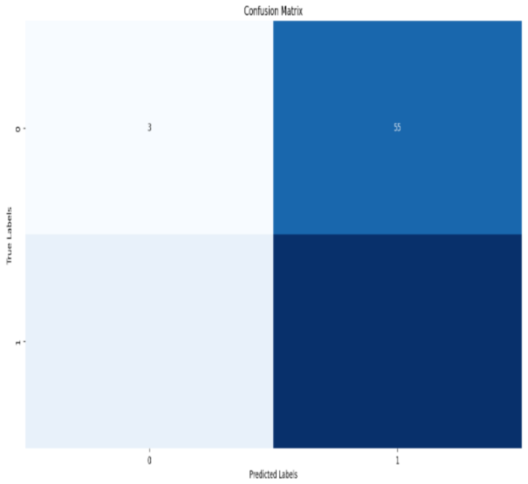

Fig 4: Confusion matrix

In accordance to this figure payments is the most prevalent business transactions type followed by cashout transfer and debit the level of success and effectiveness of forecasting fraud cash payments will be determined by the precision of the model used which is not shown in the table the real success rate must be estimated based on the success rate of the models parameters including recall precision accuracy and f1 score. the acuity is 97% the amount of precession is 90% the memory recall is 91% and the score for the f1 test is 95% however the bagging extractor performed best on an evenly matched dataset the exactness precession recall that and f1-score have all reached 97%

5. Conclusion

Good estimation outcomes can be achieved through both unbalanced and evenly distributed datasets the bagging procedure a classification method decision- tree classifier and the randomly generated forest classifier produced the highest outcomes and detecting an average of 99.50 of bogus transactions while not classifying some non-fraud transactions as fraudulent there is no perfect model and there is bound to be some level of compromise amongst both accuracy and recall its up to the organization and its aims to determine which method is best in each specific case.

References

1. Arun, K., Garg Ishan and Kaur Sanmeet. “Loan Approval Prediction based on Machine Learning Approach.” (2016).

2. G. Arutjothi and C. Senthamarai, "Prediction of loan status in commercial bank using machine learning classifier," 2017 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, pp. 416-419, 2017.

3. I Sadgali, N. Sael and F. Benabbou, “Performance of machine learning

techniques in the detection of financial frauds”, in Second International Conference on Intelligent Computing in Data Sciences 2018, Procedia Computer Science, vol. 148, pp. 45-54, 2019.

4. K. Ashuelot, A. AlGhamdi and D. P. Agrawal, "Azure ML Based Analysis and Prediction Loan Borrowers Creditworthy," 2020 3rd International Conference on Information and Computer Technologies (ICICT), 2020

5. K. R. Seeja and M. Zareapoor, “FraudMiner: A novel credit card fraud detection model based on frequent itemset mining,” Sci. World J., vol. 2014.

6. Mohammed, Emad, and Behrouz Far. “Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study.” IEEE Annals of the History of Computing, IEEE, 1 July 2018.

7. Nasr, M. H., Farrag, M. H., & Nasr, M. M. (2022). A Proposed Fraud Detection Model based on e-Payments Attributes a Case Study in Egyptian e-Payment Gateway. International Journal of Advanced Computer Science and Applications.

8. Pumsirirat, A., & Liu, Y. (2018). Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine. International Journal of advanced computer science and applications.

9. R. Rambola, P. Varshney and P. Vishwakarma, “Data Mining Techniques for Fraud Detection in Banking Sector,” 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018,pp.1-5.

10. S. K. Rumi, Ke Deng and F. D. Salim, “Crime event prediction with dynamic features”, EPJ Data Science, 2018.

11. Solving the False positives problem in fraud prediction using automated feature engineering - Wedge, Canter, Rubio et.al - October 2017

12 .Support Vector machines and malware detection - T.Singh,F.Di Troia, C.Vissagio , Mark Stamp - San Jose State University - October 2015

13. V. N. Dornadula and S. Geetha, “Credit Card Fraud Detection using Machine Learning Algorithms,” Procedia Computer. Sci., vol. 165, no. 20, pp. 631–641, 2019.

14. Xuan, Shi yang, et al. “Random Forest for Credit Card Fraud Detection.” 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), 2018.

15. Y. Lucas et al., “Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs,” Futur. Gener. Computer. Syst., vol. 102, pp. 393–402, 2020.