International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Financial Fraud Detection

1st ARUN CHAUHAN
Department of Computer Science And Engineering – Cyber Security (PIET), (Parul University). Vadodara, India.
arun.chauhan33422@paruluniversity.ac.in

2nd Ms.VRUTTI H TANDEL(A.Prof)
Department of Computer Science And Engineering – Cyber Security (PIET), (Parul University). Vadodara, India.
jafarali.kadari@techdefence.com

3rd MUKESH KUMAR
Department of Computer Science And Engineering – Cyber Security (PIET), (Parul University). Vadodara, India.

4st Sai kiran Gembali
Department of Computer Science And Engineering - Cyber Security, (PIET), (Parul University), Vadodara, India.
Saikirangembali08@gmail.com

5nd Dokra Jai Venkata Vardhan
Department of Computer Science And Engineering - Cyber Security, (PIET) (Parul University), Vadodara, India.
d.j.v.v1233@gmail.com

6rd Telukula Sahith
Department of Computer Science And Engineering - Cyber Security, (PIET) (Parul University), Vadodara, India.
sahithsahith811@gmail.com

***Abstract*—Detection of financial fraud is now a cause of major concern in the financial and banking industry because fraud techniques are becoming highly sophisticated. Classical rule- based systems are generally ineffective in detecting complex patterns of fraud, which call for more complex machine learning and artificial intelligence processes. The following paper discusses the different methodologies in detecting financial fraud, ranging from supervised and unsupervised learning to anomaly detection and deep neural network models. Also, it discusses how big data analytics and real-time monitoring of transactions contribute to making fraud detection more accurate. Key issues like biased datasets, new patterns of fraud, and AI model interpretability are also underscored in the research. Using data-driven method- ologies, financial institutions can enhance rates of fraud detection, minimize false positives, and make financial transactions more secure in general.**

***Index Terms*—component, formatting, style, styling, insert**

I. INTRODUCTION

Financial fraud is a potential danger in the banking, insurance, and e-commerce sectors, resulting in billions of dollars lost to businesses and individuals each year. Type of frauds like credit card fraud, identity theft, money laundering, and insider trading take advantage of weak points in financial systems, and the result is huge financial and reputational losses. With fraudsters evolving more complex techniques, conventional rule-based detection systems fail to keep up, producing high false positive rates and false negatives.

To overcome these issues, contemporary financial fraud detection systems utilize sophisticated technologies like machine learning, artificial intelligence (AI), and big data analytics. These methods allow organizations to detect sophisticated fraud patterns, identify anomalies in real time, and enhance the accuracy of fraud prediction models. Through the examination of large amounts of transactional data, AI-based systems can learn to evolve with new fraud schemes and minimize the use of manual investigation.

Despite all these developments, financial fraud detection is not without several challenges that it presently faces, such as data imbalances, real-time processing requirements, and interpretability of AI-driven models. This paper discusses several of the techniques applied to fraud detection, how effective they are, and with which challenges fraud detection is still grappling. By improving fraud detection processes, financial institutions can reduce losses, safeguard customers, and ensure financial system trust.

II. REVIEW

Substantial research exists on financial fraud detection owing to its significant role in preventing cyber crimes as well as from a business perspective. Some researchers have also performed literature reviews of articles that have been published in the 2000s and 2010s. To identify financial fraud, researchers mostly employ outlier detection methods (Jayakumar et.al., 2013) with highly imbalanced data sets. Various categories of financial frauds can also occur. Four types of financial fraud – financial statement fraud, transaction fraud, insurance fraud and credit fraud are suggested in one article (Jans et al., 2011). Transaction fraud in particular is the focus of this project as it relates to mobile payments. Numerous methods have been experimented with in order to identify financial fraud. Phua et al., (2004) applied Neural Networks, Na¨ıve Bayes and Decision Trees to identify automobile insurance fraud. Ravisankar et al., (2011) identify fraud in financial statements of Chinese companies, another research employed SVM, Genetic Programming, Logistic Regression and Neural Networks. Density-based clustering (Dharwa et al., 2011) and cost-sensitive Decision Trees (Sahin et al., 2013) have been applied to credit card fraud. Sorournejad et al., (2016) addresses both supervised and unsupervised machine learning-based methods encompassing ANN (Artificial Neural Networks), SVM, HMM (Hidden Markov Models), clustering. Wedge et al., (2018) discuss the issue of data imbalance that lead to an extremely large number of false positives, and certain research papers suggest methods to overcome this issue. But there is hardly any literature on the detection of fraudulent mobile payments, likely owing to comparatively recent developments in the technology.

III. METHODOLGY

This methodology served as the deliverables of the project. It describes the results of each phase that was tried out and do a comparison between them to identify which is the optimum method to solve the fraud detection issue.

IV. LITRATURE SURVEY

Financial frauds have been researched thoroughly by academic and industrial studies because they play a vital role in various critical industries. Hence, fraud detection has been a delicate issue over the last years in several areas by numerous surveys and review articles. They encompass fraud types, fraud areas, and fraud detection methods and approaches. Therefore, we examined the recent research studies and techniques to detect fraud in financial areas with the help of data mining techniques and correlate the prevailing trends There has been a wealth of published research literature that has investigated the use of anomaly detection methods in different applications, which has been the subject of interest for most survey and review articles over the past few years. Of these surveys, some have emphasized a wide range of applications, approaches, and methods that have had a remarkable influence on future research in many areas. Hodge and Austin wrote one of the earliest reviews on anomaly or outlier detection methods in 2004, giving a detailed overview on the topic (Hodge and Austin, 2004). The literature offers extensive background on outliers or anomalies and the difficulties of detecting them and a detailed review of early statistical, machine learning and ensemble approaches used to the task. In 2009, Chandola et al. also surveyed the other anomaly detection methods suggested in literature not previously included in Hodge and Austin, offering greater understanding of the other real-world applications they are utilized in (Chandola et al., 2009). In 2012, Zimek et al. released a survey published on unsupervised anomaly detection methods in particular for high-dimensional numerical data and explained the concepts of the 'curse of dimensionality' in extensive detail (Zimek, Schubert, and Kriegel, 2012). The reading included comparisons between two groups of specialized algorithms: those that deal with the issue of irrelevant features or attributes and others dealing with issues of efficiency and effectiveness (Zimek et al., 2012). Temporal data also presents another problem for anomaly detection, one that was surveyed in great detail by Gupta et al. in 2014 (Gupta, Gao, Aggarwal, and Han, 2014). With the evolution of computational powers allowing for temporal data of different types to become available, the authors comprehen- sively survey the methods that have been made possible for anomaly detection from time-series data (Gupta et al., 2014). The authors offer valuable insights into different applications of temporal anomaly detection and related challenges in each application.

V. RECOMMENDATIONS

By this project, we proved that it is possible to detect fraudulent transactions in financial transactions data with highly precise accuracy even with the high-class imbalance. We give the following suggestions from this exercise - Fraud detection in transactions data in which transaction amount and balances of the recipient and originator are known can be best executed using tree-based algorithms such as Random Forest Employing dispersion and scatter plots to represent the fractioning between fraud and transactions non-fraud selection is critical for proper feature choice To remedy high-class imbalance which is a feature of most fraud detection tasks, sampling strategies such as oversampling, undersampling, SMOTE can be utilized. Although there are restraints in computing necessities with this type of measures, particularly dealing with large datasets. When measuring performance in fraud detection systems, there must be a high level of caution on selecting the correct measure. Recall parameter is a good measure because it traps whether an adequate number of fraudulent transactions are being correctly labeled or not. We shouldn't depend on accuracy alone as it can be deceptive.

VI. ANALYSIS

We cleaned the financial transactions data and built a machine learning model to identify fraud. Data cleaning, exploratory analysis and predictive modeling were involved in the analysis. In data cleaning, we verified for missing values, changed data types and summarized variables in the data. In an exploratory analysis, we examined the class imbalance, and drilled down into each of the variables, specifically transaction type, transaction amount, balance and time step. We also found derived variables that could assist in fraud detection. We have also plotted some graphs to have a better visualization of the data and derive insights. In predictive modeling, we tried Logistic Regression and Random Forest algorithms. We found that Random Forest works best for this use case with near 100results by undersampling, but the outcome was the same due to a lot of the data being excluded. We made sure that there is no overfitting in the models using cross-validation. We can say that fraud detection in financial transactions is effective in this labeled dataset, and the best algorithm for this is Random Forest.

VII. FUTURE WORKS

Financial fraud detection is a developing subject in which it is preferable to remain ahead of the offenders. Further- more, it is clear that there exist still areas of smart fraud detection which remain unexplored. In this section we outline some of the most important problems related to financial fraud detection and propose research directions. Some of the problems identified and challenges are as follows: • Common classification issues: CI and data mining-based financial fraud

VIII. CONCLUSION

In summary, we were able to successfully create a framework for identifying fraudulent transactions in financial data. This framework will aid in understanding the nuances of fraud detection like the generation of derived variables that can potentially differentiate the classes, dealing with class imbalance and selecting the appropriate machine learning We tested

two machine learning algorithms – Logistic Regression and Random Forest. The Random Forest algorithm provided much better results than Logistic Regression tree-based algorithms perform well on welldifferentiated classes of transactions data. This also highlights the importance of performing careful exploratory analysis to know the data in detail prior to building machine learning models. From this exploratory analysis, we obtained a few features that separated the classes better than the original data.

IX. REFERENCE

1. E. Ngai et.al., The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature, Decision Support Systems. 50, 2011, 559–569 2. Albashrawi et.al., Detecting Financial Fraud Using Data Mining Techniques: A Decade Review from 2004 to 2015, Journal of Data Science 14(2016), 553-570 3. TESTIMON @ NTNU, Synthetic Financial Datasets for Fraud Detection, Kaggle, retrieved from https://www.kaggle.com/ntnu-testimon/paysim1 4. Jayakumar et.al., A New Procedure of Clustering based on Multivariate Outlier Detection. Journal of Data Science 2013; 11: 69- 84 5. Jans et.al, A Business Process Mining Application for Internal Transaction Fraud Mitigation, Expert Systems with Applications 2011; 38: 13351–13359 6. Phua et.al., Minority Report in Fraud Detection: Classification of Skewed Data. ACM SIGKDD Explorations Newsletter 2004; 6: 50-59. 7. Dharwa et.al., A Data Mining with Hybrid Approach Based Transaction Risk Score Generation Model (TRSGM) for Fraud Detection of Online Financial Transaction, International Journal of Computer Applications 2011; 16: 18-25. 8. Sahin et.al., A Cost-Sensitive Decision Tree Approach for Fraud Detection, Expert Systems with Applications 2013; 40: 5916–5923. 9. Sorournejad et.al., A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective, 2016 10. Wedge et.al., Solving the False Positives Problem in Fraud Prediction Using Automated Feature Engineering, Machine Learning and Knowledge Discovery in Databases, pp 372-388, 2018 1. C. Sullivan and E. Smith, Trade-Based Money Laundering: Risks and Regulatory Responses, pp. 6, 2012. Show in Context Google Scholar 2. Trade-Based Money Laundering Flourishing, May 2009, [online] Avail- able: http://www.upi.com/TopNews/2009/05/11/Trade-based-money-laundering-flourishing/UPI-17331242061466. Google Scholar 3. L. Akoglu, M. McGlohon and C. Faloutsos, ”OddBall: Spotting anomalies in weighted graphs”, Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining, pp. 410-421, 2010. Show in Context CrossRef Google Scholar 4. V. Chandola, A. Banerjee and V. Kumar, ”Anomaly detection: A survey”, ACM Comput. Surv., vol. 41, 2009. Show in Context CrossRef Google Scholar 5.

W. Eberle and L. Holder, ”Mining for structural anomalies in graph-based data”, Proc. DMin, pp. 376-389, 2007. Show in Context View Article Google Scholar 6. C. C. Noble and D. J. Cook, ”Graph-based anomaly detection”, Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 631-636, 2003. Show in Context CrossRef Google Scholar 7. H. Tong and C.-Y. Lin, ”Non-negative residual matrix factorization with application to graph anomaly detection”, Proc. SIAM Int. Conf. Data Mining, pp. 1-11, 2011. Show in Context CrossRef Google Scholar 8. S. Wang, J. Tang and H. Liu, ”Embedded unsupervised feature selection”, Proc. 29th AAAI Conf. Artif. Intell., pp. 470-476, 2015. Show in Context CrossRef Google Scholar 9. Z. Lin, M. Chen and Y. Ma, The Augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices, 2010, [online] Available: https://arxiv.org/abs/1009.5055. Google Scholar 10. J. Sun, H. Qu, D. Chakrabarti and C. Faloutsos, ”Neighborhood formation and anomaly detection in bipartite graphs”, Proc. 15th IEEE Int. Conf. Data Mining, pp. 8, Nov. 2005. Show in Context Google Scholar 11. A. Patcha and J.-M. Park, ”An overview of anomaly detection techniques: Existing solutions and latest technological trends”, Comput. Netw., vol. 51, no. 12, pp. 3448-3470, Aug. 2007. Show in Context CrossRef Google Scholar 12.

W. Li, V. Mahadevan and N. Vasconcelos, ”Anomaly detection and localization in crowded scenes,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 1, pp. 18-32, Jan. 2014. Show in Context View Article Google Scholar 13. K. Henderson et al., ”It’s who you know: Graph mining using recursive structural features,” Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 663-671, 2011. Show in Context CrossRef Google Scholar 14.

F. Keller, E. Mu¨ller and K. Bohm, ”HiCS: High contrast subspaces for density-based outlier ranking”, Proc. ICDE, pp. 1037-1048, Apr. 2012. Show in Context View Article Google Scholar 15.

D. Koutra, E. Papalexakis and C. Faloutsos, ”Tensorsplat: Spotting latent anomalies in time”, Proc. PCI, pp. 144-149, Oct. 2012. Show in Context View Article Google Scholar 16.

J. H. M. Janssens, I. Flesch and E. O. Postma, ”Outlier detection with one-class classifiers from ML and KDD”, Proc. ICMLA, pp. 147-153, Dec. 2009. Show in Context View Article Google Scholar 17. N. A. Heard, D. J. Weston, K. Platanioti and D. J. Hand, ”Bayesian anomaly detection methods for social networks”, Ann. Appl. Statist., vol. 4, no. 2, pp. 645-662, 2010. Show in Context CrossRef Google Scholar 18. J. Tang and H. Liu, ”CoSelect: Feature selection with instance selection for social media data”, Proc. SIAM Int. Conf. Data Mining, pp. 1-9, 2013. Show in Context CrossRef Google Scholar 19. Z. He, X. Xu and S. Deng, ”Discovering cluster-based local outliers”, Pattern Recognit. Lett., vol. 24, no. 9, pp. 1641-1650, 2003. Show in Context CrossRef Google Scholar 20. M. Gupta, J. Gao, C. C. Aggarwal and J. Han, Outlier Detection for Temporal Data, San Rafael, CA, USA:Morgan Claypool, 2014. Show in Context Google Scholar 21. J. Tang, Y. Chang and H. Liu, ”Mining social media with social theories: A survey”, ACM SIGKDD Explorations Newslett., vol. 15, no. 2, pp. 20-29, 2013. Show in Context CrossRef Google Scholar 22. I. S. Dhillon, S. Mallela and D. S. Modha, ”Information-theoretic co-clustering”, Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 89-

98, 2003. Show in Context CrossRef Google Scholar 23. Q. Gu and J. Zhou, ”Co-clustering on manifolds”, Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 359-368, 2009. Show in Context CrossRef Google Scholar 24. K. Sim, V. Gopalkrishnan, A. Zimek and G. Cong, ”A survey on enhanced subspace clustering”, Data Mining Knowl. Discovery, vol. 26, no. 2, pp. 332-397, 2013. Show in Context CrossRef Google Scholar 25. S. Mcskimming, ”Trade-based money laundering: Responding to an emerging threat”, Deakin Law Rev., vol. 15, no. 1, 2010. Show in Context CrossRef Google Scholar 26. E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen and X. Sun, ”The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature”, Decision Support Syst., vol. 50, no. 3, pp. 559-569, 2011. Show in Context CrossRef Google Scholar 27.

A. Srivastava, A. Kundu, S. Sural and A. K. Majumdar, ”Credit card fraud detection using hidden Markov model,” IEEE Trans. Depend. Sec. Comput., vol. 5, no. 1, pp. 37-48, Jan./Mar. 2008.