FRAUD DETECTION WITH NATURAL LANGUAGE PROCESSING

[1]MUKTHA RAMESH KUMAR, [2]Dr. MANISH VASHNEY

[1]Research Scholar, Maharishi University of Information Technology, Lucknow, U.P. 226013, INDIA.
[2]Research Supervisor, Maharishi University of Information Technology, Lucknow, U.P. 226013, INDIA

KEYWORDS
Natural language processing, Fraud detection,Varying class imbalance, Feature engineering,E-banking.

ABSTRACT
Automated fraud detection may aid companies in protecting user accounts, a job that is particularly difficult owing to the scarcity of proven fraudulent transactions. A significant portion of the existing literature primarily addresses credit card theft while neglecting the emerging field of internet banking. Nevertheless, there is a dearth of readily accessible data for both. The absence of readily accessible data impedes the advancement of the field and restricts the exploration of possible remedies. This work accomplishes three main objectives. Firstly, we present FraudNLP, which is the initial anatomized dataset accessible to the public for online fraud detection. Secondly, we evaluate various machine and deep learning techniques using multiple assessment metrics. Lastly, we demonstrate that online actions adhere to patterns similar to natural language, making them amenable to successful analysis using natural language processing methods.

1 INTRODUCTION

Fraud detection systems assess every transaction to find any instances of fraud. Subsequently, a bank employee may evaluate the level of danger based on their predetermined safety limit and make a decision, such as blocking the transaction or requesting more information. Given the broad adoption and high volume of transactions in Internet and mobile banking, it is unfeasible for a person to manually monitor and identify all instances. Preventing and promptly identifying fraud is essential for establishing and maintaining client trust in these platforms, as well as for the bank to prevent charge-backs. Automated fraud detection may identify possibly fraudulent transactions, serving as the last safeguard before an employee intervenes to resolve the issue.

This study addresses the issue of fraud detection specifically in the context of online and mobile banking transactions. Consistent with the prevailing method in previous research, we originally tackled the challenge as an unbalanced binary classification problem. Transactions were represented as isolated occurrences, disregarding any preceding user activities. In addition, we dedicated significant time and effort to performing the intricate process of feature engineering, which involves extracting

regularly used user profiling characteristics. We consolidated all the API calls made during user sessions that included a transaction into a single sequence of actions. This sequence concluded with the transaction, either a transfer or a payment. We then classified each sequence as fraudulent or not based on the Bank's determination of the final transaction's legitimacy. In this context, we analyse the series of user interactions, such as logging in, logging out, checking the account balance, and reviewing recent transaction history, that culminate in a specific kind of transaction, such as a generic transfer, rent or bill payment, or quick payment. We use the whole sequence of activities, in addition to other relevant characteristics, to assess the transaction. By extracting sequences in this manner, we were able to: (a) frame fraud detection as a problem of classifying sequences, which minimised the need for extensive feature engineering, and (b) become the first to make our dataset publicly available.

Fraud detection may be used in two distinct settings: online and offline. The former pertains to real-time detection, which is supposed to operate in a proactive manner, alerting an employee to intervene. This situation is more effectively handled by high-precision algorithms. The latter pertains to the assessment of past data to identify any potential instances of undetected fraud. This issue is more effectively resolved with high-recall approaches. Unlike the majority of published research that ignores this insight, we assess all of our approaches in both scenarios. Our dataset experimentation shows that online behaviours have parallels with natural language. Additionally, NLP-based features may enhance the performance of fraud classifiers, surpassing current approaches. These features need less programming and prioritise privacy.

Overall, our contributions are the following:
FraudNLP is a novel dataset for fraud detection that is publicly accessible and anonymous. It comprises 105,303 transactions and is derived from the behaviours of 2,000 individuals prior to the transactions.
We evaluate the performance of machine and deep learning algorithms on our dataset, taking into account both online and offline detection. These aspects, while crucial, have been neglected in previous research.
We demonstrate that using privacy-preserving natural language processing (NLP) features enhances the efficacy of machine learning and surpasses the current cutting-edge methods.
Through evaluating our most effective classifier on various class imbalance configurations, we demonstrate that the difficulty of the challenge significantly increases as the imbalance becomes greater, which aligns more closely with the actual nature of the work.

The subsequent section of the essay first examines previous research and then introduces the novel dataset. Subsequently, an empirical analysis and a section for debate are presented. This research is concluded with a summary of our findings.

2 Related Work

In their 2015 study, Michele Carminati and colleagues introduced BankSealer, a system that employs a semi-supervised methodology to assess the level of suspicion associated with user transactions and score them accordingly. The primary approach they use is anomaly detection techniques to construct personalised behavioural profiles for users based on their transaction history, without using any sequential data. In a separate publication, Michele Carminati and colleagues (2018) introduce a framework named FraudBuster, which aims to identify instances of financial fraud that include gradually embezzling tiny sums of money. Their system utilises a model to represent the user's spending behaviour over a period of time and identifies fraudulent transactions as those that depart from the established model and alter the user's spending profile.

Kovach and Ruggiero (2011) provide a method that generates a risk score by merging changes in behaviour at both the individual (user) level and the collective level, including all users in the bank. Similarly, they consider transactions as isolated moments in time without any preceding action history and incorporate contextual details through meticulously designed statistical features (such as differential analysis to measure abnormality at a local level, a probabilistic model at a global level, and the Dempster-Shafer theory to combine the two). In addition, users are required to download a separate programme to enable device fingerprinting, which adds complexity to the implementation of this approach, particularly for those who prioritise privacy.

In contrast to previous research that only examines transaction sequences (Wang 2021; Forough and Momtazi 2022, 2021), our approach involves analysing the sequence of user activities that occur before a transaction to assess its legality. By formulating the issue in this way, we were able to treat fraud detection as a task of classifying sequences, minimising the need for extensive feature engineering. The findings of our study demonstrate that the suggested technique of feature engineering produces effective solutions, and even outperforms the current best method when paired with basic anomaly detection features. Our characteristics include inherent anonymity. Therefore, via this endeavour, we are making the first dataset for online fraud detection publicly accessible.

Recurrent Neural Networks have recently been used to tackle the job of fraud detection. These networks are capable of extracting information from the past card transaction records of each user (Branco et al., 2020; Roy et al., 2018). Achituve et al. (2019) consider the past transactions of each user as a sequential series. By using attention-based recurrent neural networks (RNNs), they achieve enhanced performance and attention scores, which are used to provide interpret-ability to the output of their classifier. The information of each transaction, including the day of the week, hour of the day, amount, and device identifier, is encoded into several variables. Subsequently, the process of learning embedding vectors is initiated. The transactions

were grouped into sequences, allowing for the use of historical information of varying breadth for each customer. The findings of Jurgovsky et al. (2018) demonstrate that LSTM enhances the accuracy of detecting fraudulent transactions in an offline setting, as compared to their baseline random forest classifier. Moreover, the integration of sequential and non-sequential learning approaches has the potential to further enhance the effectiveness of fraud detection. In 2018, Kunlin introduced a new algorithm called FraudMemory for detecting fraud. This algorithm used advanced ways for representing features in order to accurately display people and logs with various kinds in financial systems. The model effectively captures the sequential patterns of each transaction and utilises memory networks to enhance performance. FraudMemory's flexibility to idea drift in shifting situations was improved by integrating memory components.

Forough and Momtazi (2022) introduced a credit card fraud detection model in the field of sequence classification, using deep neural networks and probabilistic graphical models. The research conducted a comparison between their model and the baseline using real-world datasets. It was discovered that taking into account the underlying sequential connections of transactions and anticipated labels led to better outcomes. Additionally, a unique undersampling algorithm was presented and shown favourable outcomes in comparison to other oversampling and undersampling techniques. Similarly, in the realm of behavioural modelling for fraud detection, Wang (2021) and Rodríguez et al. (2022) closely align with us in terms of their philosophy and range of features. Their research demonstrates that detecting fraud in online payment systems does not necessarily require the identification of unauthorised behaviour. Their suggested solution is an account risk prediction technique that aims to anticipate fraud by analysing a user's previous transaction sequence.

3 Dataset:

The FraudNLP dataset discussed in this study1 pertains to individuals who engaged with a European bank via its online and mobile banking systems.Two Firstly, we will cover the process of developing the dataset. Next, we examine the numerical characteristics that are often derived in previous studies (Baesens et al., 2021; Wedge et al., 2019), as well as the factors associated with anomaly detection that have been shown to enhance the accuracy of detection (Baesens et al., 2021). The sequential data introduced in this study are displayed at the end.

We analysed transactions occurring from February 1st to October 31st, 2020, including all instances of fraud as well as a substantial number of valid transactions. Amongst the documented transactions during this time, a total of 101 were identified and confirmed by a bank employee as fraudulent. Initially, we randomly picked 10,000 individuals from the bank servers who had no instances of fraudulent transactions during the course of nine months. Due to insufficient user engagement with the bank services, we had to exclude their logs since they did not provide significant information on their spending patterns. A minimum threshold of 12

transactions was used to filter the transactions, resulting in a final dataset that includes logs from 2,000 people. The logs of 97 individuals were subsequently included, namely those users who had at least one transaction verified as fraudulent by a bank staff during this period. It should be noted that this approach results in an underestimating of the actual proportion of individuals engaging in fraudulent transactions. This is because the 2,000 people we analysed represent just a tiny subset of all users who did not have any recorded fraudulent behaviour during the specified time period. There is a limited number of fraudulent transactions, with just 101 cases out of a total of 105,303 transactions (0.096%). It is important to acknowledge that the bank may not detect all fraudulent transactions. While the confirmed fraud instances in our dataset are reliable, there may be unidentified and unreported occurrences.

3.1 Recency, frequency, monetary

The predominant characteristics used in fraud detection are numerical features that characterise the user's behaviour according to the Recency, Frequency, Monetary (RFM) concept, as described in Baesens et al. (2021). After studying their work, we include elements that conform to it. A frequency table was produced by starting with the earliest transaction accessible per user, which is 9 to 11 months in this dataset, and then increasing it. After a one-month period of first use, this table records the individuals and organisations that each customer often interacts with, including both one-time and regular payments such as rent, subscriptions, or mortgage payments. Additionally, the natural logarithm of each transaction value (in EUR) was obtained as per the client's request and the corresponding server answer. These values were standardised by calculating the mean and standard deviation of the quantities in the training set. The requests and answers included beneficiary information, which was used to generate the growing relative frequency tables for each beneficiary..

Table 1: The features are derived from action sequences (top), RFM (centre), and anomaly detection.

Parameter	Type	Dimensions
User action sequence	Integers	128
Time between actions (ms)	Integers	128
Log of transaction amount	Float	2
Time to execute transaction (ms)	Float	2
Device frequency	Float	2
IP address frequency	Float	2
Application frequency	Float	2

A training subset consisting of 100 trees was established in the forest, and the anticipated anomaly scores were obtained for each training observation.

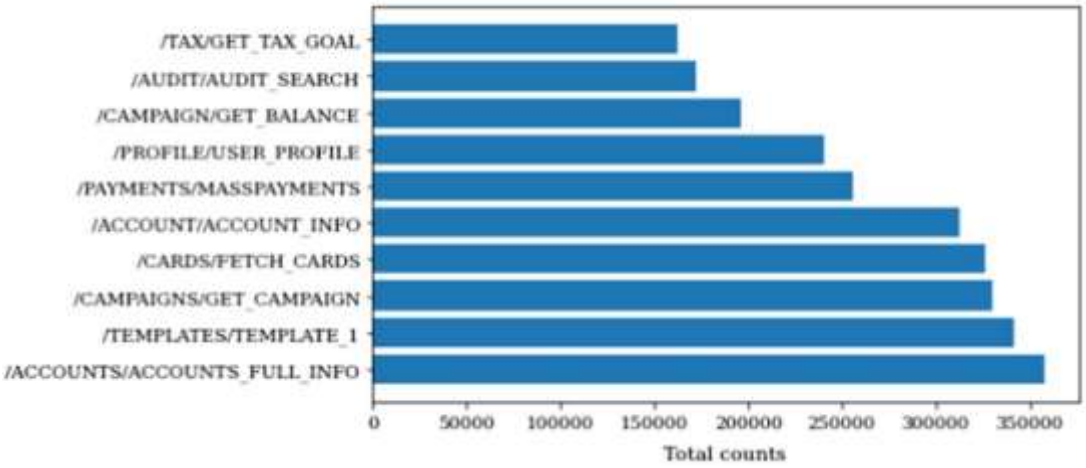

Fig. 1: The activities that occur most often in the dataset

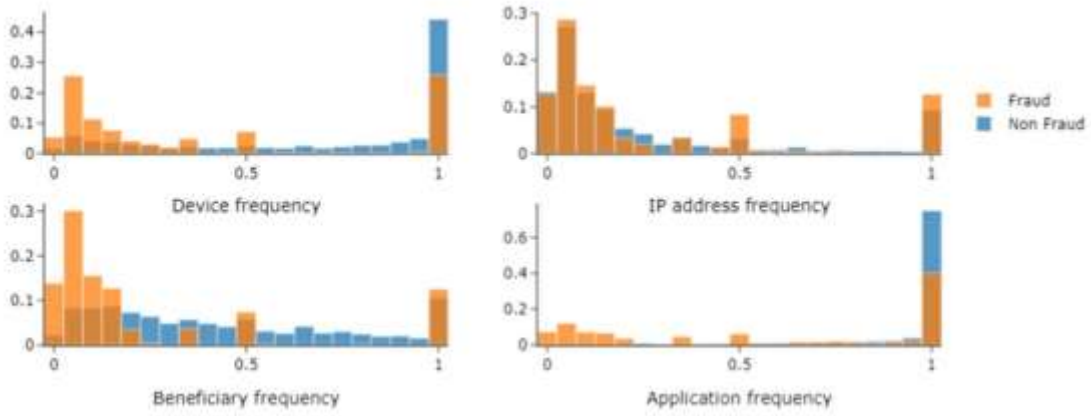

Fig. 2: The histogram displays the frequency of the numerical characteristic on the X-axis for both fraudulent (orange) and non-fraudulent (blue) transactions. (Colour figure available online)

Table 2: The distribution of classes in the training, validation, and test subsets of our dataset is shown, showing the absolute number and percentage for each class. The occurrence of fraudulent cases is below 0.1% in all subgroups.

	Transaction label	
	Normal (%)	Fraudulent (%)
Training	64,268 (98.918)	61 (0.097)
Validation	20,170 (98.916)	21 (0.102)
Test	20,069 (98.925)	21 (0.103)

3.2 Exploratory analysis

The instances of fraudulent actions (highlighted in orange) are few and dispersed among the non-fraudulent actions (highlighted in blue), in terms of the time intervals

between consecutive actions within a single sequence. When examining the activities inside the sequences, we see that the five most common actions are likewise the most common within each class. An interesting anomaly to note is the URL '/loans/list', which ranks as the 6th most common in fraudulent sequences, but drops to the 7th most common in the far larger number of non-fraudulent sequences. Additionally, it is worth mentioning that the endpoint '/card/fetchcards' is much more prevalent in fraudulent transactions, ranking 10th, as opposed to non-fraudulent transactions, where it ranks 21st.

Fig. 3 displays the distribution of each numeric attribute according to transaction status, which is another noteworthy statistic. It becomes clear that our characteristics reveal the distinct behaviour that we anticipated fraudulent transactions to exhibit. Within the device frequency histogram, we see two clearly defined regions where fraudulent activity takes place: the regions characterised by very high and extremely low frequencies. The former may be ascribed to device theft, in which the same device is used to perpetrate fraud, while the latter can be ascribed to account takeover, in which a new device is utilised. However, when examining the IP address frequency histogram, there is little disparity in the distribution of lawful and unlawful activities, with the exception of the medium and high-frequency regions. This is likely due to the fact that in both account takeover and device theft situations, a different IP address is used, resulting in the absence of the two separate sections seen in the other features.

While the statistics of the sequences themselves did not uncover many intriguing patterns, it is anticipated that there would be more inconsistencies in action sub-sequences. This is because fraudsters could use the same activities to avoid detection, but in a different sequence, such as viewing the available cards. When analysing the occurrence of action trigrams (sequences of three consecutive actions), we see that the most common case frequently varies between the two groups, when we consider non-fraudulent samples of similar size. After doing one thousand iterations of this sampling process, we can confidently state that this finding is statistically significant, with a P-value of 0.02.

4 Experiment analysis

For all of our trials, we used a stratified split of 60% for training, 20% for development, and 20% for testing. Additionally, we conducted Monte Carlo 5-fold Cross-Validation. We conducted training and evaluation of four machine learning classifiers using our dataset, with the objective of predicting the fraudulent nature of a transaction. We used Logistic Regression (LR), Random Forests (RF), k closest neighbours (kNN), and Support Vector Machines (SVM).

The significant class disparity exacerbates the difficulty of the assignment. For the assessment, we used evaluation criteria that are unaffected by the skewed character of the situation. Specifically, we used the F1 score and the Area Under the Precision-Recall Curve (AUPRC). We selected the second option, the area under the ROC

curves (ROC-AUC) (Saito and Rehmsmeier 2015), since it is less affected by differences in class distribution. A majority classifier in this job would have a precision of 0.096%, which would be appropriately represented by the AUPRC (0.096%), in contrast to the ROC-AUC (50%).

The area characterised by both high precision and strong recall. This presents a dilemma since instances of fraud are often identified either immediately (online) when accuracy is crucial, or offline when completeness is crucial. Therefore, we recommend evaluating approaches using the high-precision F05 metric for the online option and the high-recall F2 metric for the offline configuration. We fine-tuned the classification threshold on the development set for each fold, for all three F-scores.

Table 2: Evaluation of machine and deep learning fraud classifiers

Model	Accuracy
LR	92%
RF	93.25%
KNN	98.86%
SVM	96.23%
LSTM	95.26%
CNN	98.23%
TCN	97.23%

5 Conclusion

The findings of our study indicate that using a different, time-based strategy to data engineering may use machine learning techniques to achieve superior performance compared to conventional feature engineering methods. Confirming the hypothesis of Baesens et al. (2021), it has been shown that using more intricate techniques, such as dense representations and deeper neural networks, does not always provide better results compared to conventional machine learning methods. When we only use TF-IDF features from our transactional corpus, excluding RFM, or solely rely on conventional RFM-based features, excluding TF-IDF, the performance significantly decreases when using standard classification measures. We further contended that assessment should include two specific situations in this task: the online scenario, which necessitates quick alerting; and the offline scenario, when Recall is at risk (i.e., spotting potentially overlooked occurrences). We proposed the use of F0.5 and F2 as benchmarks for evaluating our models.

Our research presents a dataset that has been made anonymous and accessible to the public for the purpose of detecting online fraud. We showcase the effectiveness of natural language processing (NLP) techniques in analysing online behaviours and getting the best possible outcomes, all while ensuring the privacy of the users.

Ultimately, we noticed that the presentation of findings in academic publications often relies on various presumed scenarios of imbalance. This undermines the capacity to make meaningful comparisons and impedes the advancement of the field. In order to tackle this issue, we chose to use a fluctuating imbalance curve. This approach not only enabled us to examine the level of difficulty for different imbalance settings, revealing that larger imbalance intensifies the challenge, but also allowed a comparison with the findings documented in previous published research. By using the same machine learning technique and a reduced number of conventional features in comparison to Baesens et al. (2021), the incorporation of NLP-based characteristics resulted in a substantial improvement in performance. This enhancement was achieved without incurring any more expenses, but rather by using the advantage of privacy, since we make our data available for public use.

References

[1] Achituve, I., Kraus, S., & Goldberger, J. (2019) Interpretable online banking fraud detection based on hierarchical attention mechanism. In 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP) (pp. 1–6). IEEE.

[2] Baesens, B., Höppner, S., & Verdonck, T. (2021). Data engineering for fraud detection. Decision Support Systems. https://doi.org/10.1016/j.dss.2021.113492

[3] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5, 135–146.

[4] Branco, B., Abreu, P., Gomes, A. S., Almeida, M. S. C., Ascensão, J. T., & Bizarro, P. (2020). Interleaved sequence RNNS for fraud detection. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ttps://doi.org/10.1145/3394486.3403361

[5] Carminati, M., Baggio, A., Maggi, F., Spagnolini, U., & Zanero, S (2018) FraudBuster: Temporal Analysis and Detection of Advanced [1] Financial Frauds, pp. 211–233. https://doi.org/10.1007/978-3-319-93411-2_10

[6] Carminati, M., Caron, R., Maggi, F., Epifani, I., & Zanero, S. (2015). Banksealer: A decision support system for online banking fraud analysis and investigation. Computers & Security, 53, 175–186.

[7] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: synthetic minority oversampling technique. Journal of Artifcial Intelligence Research, 16, 321–357.

[8] Fawcett, T., & Provost, F. (1997). Adaptive fraud detection. Data Mining and Knowledge Discovery, 1, 291–316. https://doi.org/10.1023/A:1009700419189

[9] Forough, J., & Momtazi, S. (2021). Ensemble of deep sequential models for credit card fraud detection. Applied Soft Computing, 99, 106883. https://doi.org/10.1016/j.asoc.2020.106883

[10] Forough, J., & Momtazi, S. (2022). Sequential credit card fraud detection: A joint deep neural network and probabilistic graphical model approach. Expert Systems, 39(1), 12795. https://doi.org/10.1111/exsy.12795

[11] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
[12] Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., & Caelen, O. (2018). Sequence classifcation for credit-card fraud detection. Expert Systems with Applications,100, 234–245. https://doi.org/10.1016/j.eswa.2018.01.037
[13] Kovach, S., & Ruggiero, W. V. (2011). Online banking fraud detection based on local and global behavior. In Proc. of the Fifth International Conference on Digital Society, Guadeloupe, France (pp. 166–171).
[14] Kunlin, Y. (2018). A memory-enhanced framework for fnancial fraud detection. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 871–874). https://doi.org/10.1109/ICMLA.2018.00140
[1] LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. In The Handbook of Brain Theory and Neural Networks, (Vol. 3361(10)).
[15] Liu, F. T., Ting, K., & Zhou, Z.-H. (2009). Isolation forest. In 2008 8th IEEE International Conference on Data Mining (pp. 413–422). https://doi.org/10.1109/ICDM.2008.17.
[16] Prasadu Peddi, & Dr. Akash Saxena. (2016). STUDYING DATA MINING TOOLS AND TECHNIQUES FOR PREDICTING STUDENT PERFORMANCE. *International Journal of Advance Research and Innovative Ideas In Education*, 2(2), 1959-1967.
[17] Mehana, A., & Nuci, K. P. (2020) Fraud Detection using Data-Driven Approach.Nguyen, T.T., Tahir, H., Abdelrazek, M., & Babar, A. (2020). Deep Learning Methods for Credit Card Fraud Detection.
[18] Panigrahi, S., Kundu, A., Sural, S., Majumdar, A.K., et al. (2009). Credit card fraud detection: A fusion approach using dempster-Shafer theory and Bayesian learning. Information Fusion, 10(4), 354–363. https://doi.org/10.1016/j.infus.2008.04.001. Special Issue on Information Fusion in Computer Security.
[19] Patel, Y., Ouazzane, K., Vassilev, V., & Li, J. (2019). Remote banking fraud detection framework using sequence learners. Journal of Internet Banking and Commerce, 24(1), 1–31.
[20] Rinku, Narang, S. K., & Kishore, N. (2023). Issues in Credit Card Transactional Data Stream: A Rational Review. Lecture Notes in Networks and Systems (Vol. 421, pp. 775–789). www.scopus.com
[21] Rodríguez, J. F., Papale, M., Carminati, M., & Zanero, S. (2022). A natural language processing approach for fnancial fraud detection. In Proceedings of the Italian Conference on Cybersecurity ITASEC 2022, Rome, Italy, June 20–23, 2022 (Vol. 3260, pp. 135–149). CEUR-WS.org.
[22] Prasadi Peddi and Akash Saxena (2014) "EXPLORING THE IMPACT OF DATA MINING AND MACHINE LEARNING ON STUDENT PERFORMANCE", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.1, Issue 6, page no.314-318, November-2014, Available :http://www.jetir.org/papers/JETIR1701B47.pdf

[23] Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifers on imbalanced datasets. PloS one, 10(3), 0118432–0118432. https://doi.org/10.1371/journal.pone.0118432
[24] Wang, C. (2021). The behavioral sign of account theft: Realizing online payment fraud alert. In Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artifcial Intelligence (pp. 4511–4618).
[25] Wang, S., Liu, C., Gao, X., Qu, H., & Xu, W. (2017). Session-based fraud detection in online e-commerce transactions using recurrent neural networks. In Y. Altun, K. Das, T. Mielikäinen, D. [1] Malerba, J. Stefanowski, J. Read, M. Žitnik, M. Ceci, & S. Džeroski (Eds.), Machine Learning and Knowledge Discovery in Databases (pp. 241–252). Cham: Springer.
[26] Wedge, R., Kanter, J., Veeramachaneni, K., Moral, S., & Iglesias Pérez, S. (2019). Solving the false positives problem in fraud prediction using automated feature Engineering: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018. Proceedings, Part III, 372–388. https://doi.org/10.1007/978-3-030-10997-4_23
[27] Prasadi Peddi and Dr. Akash Saxena (2015), "The Adoption of a Big Data and Extensive Multi-Labled Gradient Boosting System for Student Activity Analysis", International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 3, Issue 7, pp:68-73.
[28]Zamini, M., & Montazer, G. (2018). Credit card fraud detection using autoencoder based clustering. In 2018 9th International Symposium on Telecommunications (IST), pp. 486–491. https://doi.org/10.1109/ISTEL.2018.8661129
[29] Zhang, Z., Chen, L., Liu, Q., & Wang, P. (2020). A fraud detection method for low-frequency transaction. IEEE Access, 8, 25210–25220. (Cited By :10)