[1]Jwalant Babubhai Baria

[2]Vanrajkumar Dineshkumar Baria

[3]Salman Yakub Bhimla

[4]Rinkalben Prajapati

[5]Mayureeben Rathva

[6]Shreyas Patel

Deep Learning based Improved Strategy for Credit Card Fraud Detection using Linear Regression

Abstract: - Credit card fraud detection is a critical challenge in the financial industry, necessitating robust and accurate methods to identify fraudulent transactions. Traditional machine learning approaches have demonstrated some success but often struggle with the dynamic and evolving nature of fraudulent activities. This paper proposes an improved strategy for credit card fraud detection by integrating deep learning techniques with linear regression models.The proposed method leverages the strengths of deep learning in capturing complex, non-linear relationships and high-dimensional patterns within transaction data, while utilizing linear regression to ensure interpretability and simplicity in the final decision-making process. Our hybrid model first employs a deep learning architecture, specifically a convolutional neural network (CNN) or a recurrent neural network (RNN), to extract meaningful features from raw transaction data. These features are then fed into a linear regression model that performs the final classification.The integration of deep learning and linear regression not only boosts performance but also provides insights into the contributing factors of fraudulent transactions, aiding financial institutions in their ongoing efforts to combat credit card fraud.

Keywords: Credit Card Fraud Detection, Deep Learning, Linear Regression, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN)

Introduction

Consumer trust is eroded and businesses suffer huge losses due to the ever-present issue of credit card theft. The need of reliable fraud detection systems has grown in recent years due to the dramatic increase in fraudulent activities associated with the widespread use of online transactions and digital payments. Quickly identifying fraudulent transactions is just half the battle; the other half is reducing the number of false positives that might derail legal transactions and annoy consumers. For a long time, credit card fraud detection has relied on old-

[1] Assistant Professor, Computer Engineering Department, Government Engineering College, Dahod

jwalant.baria@gmail.com

[2]Assistant Professor, Computer Engineering Department, Government Engineering College, Dahod

vdbaria.ce@gmail.com

[3]Assistant Professor, Computer Engineering Department, Government Engineering College, Dahod

sbhimla.comp@gmail.com

[4]Assistant Professor, Computer Engineering Department, Government Engineering College, Modasa

rinkal.prajapati@gecmodasa.ac.in

[5]Assistant Professor, Computer Engineering Department, Government Engineering College, Modasa

mayurirathva@gmail.com

[6]Assistant Professor, Computer Engineering Department, Government Engineering College Modasa

shreyas.patel@gecmodasa.ac.in

school machine learning methods like logistic regression, decision trees, and support vector machines. These techniques find patterns of suspicious activity in past transactions. Even while they've had some success, these methods can't keep up with the ever-changing strategies used by fraudsters. These models aren't dynamic enough to handle the multi-dimensional patterns and intricate, non-linear interactions seen in real-world transaction data.

The ability to automatically extract complex characteristics from massive datasets is what has made deep learning, a subset of machine learning, so revolutionary in many domains. In the fields of picture identification, NLP, and time series analysis, techniques like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have shown outstanding performance. CNNs excel at detecting local patterns and spatial hierarchies, which makes them a good fit for handling structured transaction data. However, RNNs are great with sequential data, which is useful for simulating the ebb and flow of transaction sequences over time.

Without relying heavily on feature engineering, deep learning can learn representations straight from raw data, which is its main power. Due to the nuanced and context-dependent nature of fraudulent transactions, this feature is especially useful for detecting credit card fraud. Improved fraud detection systems that can learn to spot previously unseen patterns of fraud are within reach with the help of deep learning.

The notion known as "digitization" has a significant impact on today's youth. All areas of the economy, including banking, finance, and insurance, are benefiting greatly from this digitization knowledge. The Indian banking industry should prioritise digitization because of the crucial role it plays in financial inclusion, which is primarily essential for providing clients with top-notch services and the opportunity to earn more in the future [4]. When it comes to moving money around, most people are already acquainted with the standard: internet banking[5]. Online banking is becoming more popular, which means more and more people are using it to pay for things like insurance premiums, train and bus tickets, utility bills (including power, water, and property taxes), online shopping, and more [6][8]. The efficiency of online banking is constantly improving. The rise in fraudulent activities is one big negative aspect of this growth [6]. A relatively new phenomenon, online banking has grown rapidly in recent years[7]. It is also known as e-banking or internet banking. Online banking has become an indispensable service even for the average person in recent times[8]. One innovative banking option that customers may take advantage of is electronic banking, which lets them manage their money online. Many financial services are made available via electronic banking, including those that include automated teller machines (ATMs), electronic funds transfers (EFTs), direct deposits, automatic bill payment (ABPs), and many more [9]. The financial association also stands to benefit more from this approach. Nevertheless, this banking approach offers bespoke convenience and flexibility at a lower cost than the conventional way. There were a lot of obstacles to overcome in this upgraded online banking system due to the dangers and assaults of fraud data settling [10]. These days, hackers may breach the security of online banking in a variety of ways. There has been a meteoric rise in the number of online application providers catering to both B2B and B2C markets, and with it, the number of fraudulent assaults and activities. Strong authentication mechanisms are crucial for online transactions in this setting. The need for a plethora of cutting-edge methods to improve the security of cryptographic systems has grown in tandem with the proliferation of various forms of electronic communication and information[11].

Security in Digital Banking

One of the biggest problems with online banking is keeping customer information safe. The first stages of a thief's plan to steal someone's money are fraught with difficulty and danger. However, identity thieves may easily sneak into an online bank account and take money if they have the customer's certain personal information. Physical, geographical, informational, and communicative privacy are the four basic types of privacy [12]. Information privacy, or the capacity of a person to govern one's own information, is essentially what we mean when we talk about Internet privacy. Furthermore, people cannot maintain a high level of control over their personal information and how it is used, which poses a threat to their privacy. In addition, there are three main parts to information security: availability, integrity, and secrecy. When evaluating the safety of an online store's computer systems, most people use the Certified Internal Auditor (CIA) standard [13]. The three pillars of security are susceptible to intentional human actions, technological issues, natural disasters,

unintentional occurrences, and so on. Confidentiality is the practice of limiting access to information and data to authorised users while preventing unauthorised users from gaining access. A further definition of secrecy is a promise to disclose information exclusively with approved parties [14]. Passwords and user IDs are examples of authentication procedures that are used to identify users and assist accomplish the goal of secrecy. In addition, other control methods are used to ensure confidentiality, such as restricting access to data system resources for every identifiable user. Along these lines, protecting sensitive information from threats like spam, malware, spyware, and other online threats is of the utmost importance [14]. Due to the sheer volume of banking operations, security concerns often rank high on the list of priorities for industrial bank management. In addition, by calculating many elements, banking security is guaranteed. On the other hand, bank security for businesses is a complex framework that involves several performances, such as the organization of resources within the context of operational risks, market conditions, and credit. The risk of failure due to external factors or internal processes is what process security is primarily concerned with, and it is a persistent aspect of working risk [15][16]. The safety of funds kept in automated teller machines and bank offices is related to physical security. Every internal and external operation is involved in the system's security, which is achieved via an information system [17]. Banks' primary characteristic—the safety of client deposits—has a significant impact on the retention, attrition, or gain of consumers. Hence, as a crucial part of their business, commercial banks must take several precautions to guarantee the appropriate and efficient security of their customers' deposits [16]. Remote access to bank accounts is made possible by electronic banking, which is part of the body of information measures [18]. The term "electronic banking" refers to a system that allows customers to access their bank accounts and transaction history via computers, mobile phones, and landlines. Electronic devices are used to process card payments; this is done at the customer's request and does not need the physical presence of either party, but rather the transmission of data. Customers use gadgets optimised for electronic processing for both receiving and transferring. A further feature of this electronic processing approach is the transmission of data over communication networks. Depending on the kind of contract entered into between the client and the bank, the customer may engage in either active or passive activities [17]. Every organization's first priority should be safety. Data security, both during storage and transmission, is a major concern for many groups. Due to the authorised remote access to sensitive information, there are security concerns with electronic banking and electronic commerce [10]. To meet the security requirements of online banking, banks often use a variety of security measures. Data communication between the bank and the consumer is guaranteed by the use of many mechanisms, including digital signatures, encryption, authentication procedures, and Secure Socket layer (SSL). Several security risks, including the fast-flux service network, pharming, spear-phishing, and whaling, are affecting authentication methods used by online banks. The financial sector has been in the forefront of implementing new authentication methods, including software-based solutions, e-signature technology, one-time passwords, and smart card technology [11]. To take security to the next level, new methods have emerged that use biometrics and numbers like PINs, CVVs, and address verification systems. AVS verifies the customer's address and zip code, whereas PIN and CVM verify the customer's entered numerical code. Signature and fingerprint verification are further examples of biometrics. The e-banking system makes use of rule-based procedures in addition to geographical regions and consumer control registers (both positive and negative).

Linear Regression in Fraud Detection

Although deep learning models are highly predictive, they are more often than not just "black boxes," with no transparency into how they arrive at their decisions. In financial applications, where explain ability and transparency are paramount, this lack of interpretability might be a major downside. Conversely, the interpretability of linear regression models is built right in, so it's easy to see how each input characteristic affected the final prediction. Our goal in combining deep learning with linear regression is to get the most out of both worlds: deep learning's great accuracy and flexibility and linear regression's simplicity and interpretability[13]. Financial institutions benefit greatly from this hybrid strategy since it makes the fraud detection system both effective and transparent in its ability to identify fraudulent transactions..

Research Methodology

Banks have suffered enormous losses due to the lack of an efficient detection system, making credit card fraud detection an essential procedure. Minimising losses and ensuring client satisfaction both need an efficient detection model. Credit card fraud detection may be greatly improved with the three contributions discussed in

this thesis, which build upon the domain research[14]. The first step in writing this thesis is to conduct a survey and analysis. Surveying and studying the realm of credit and debit card transactions is what this profession is all about. Also included is a review of the literature on the subject of credit and debit card transactions, including its benefits and drawbacks. A review of the current approaches reveals that batch processing is the primary use case for the models. Because of this, they aren't good candidates for processing in real time[15]. In addition, undersampling is a way that some methods deal with imbalance, while others don't seem to care about it at all. Unfortunately, this leads to data loss, which is obviously bad for getting useful findings. This paper also presents and discusses the domain's breadth and future research goals. The first piece of work suggests an ensemble model that uses stacking to effectively deal with data that is uneven. Finding the decision signatures in the complicated credit card transaction data requires identifying the stacking model. A very efficient model is produced by the rule-based combiner that is used as the aggregator mechanism; this mechanism also does some analysis. To further enhance the prediction quality, the second contribution suggests a tree-based ensemble model with bagging[16]. The model combines bagging and boosting, two ensemble methods. The bagging technique is used to merge many boosted models. Both bagging and boosted models aim to decrease model variance, but bagging also tends to decrease prediction bias. Consequently, a model for fraud detection that is both accurate and efficient is produced. The findings show that the procedure was efficient, with good performance levels. Finally, a heterogeneous ensemble is suggested as a solution to deal with data complexity. In order to provide findings in real-time, the study employs a tree-based meta classifier, the goal of which is to decrease prediction time. Time is still a big limitation, even if the prior contribution produced useful outcomes. The last contribution resolves this issue. There are two tiers of prediction that the model makes. The tree-based bagging model Random Forest is used for the first level of prediction, while a mixture of Decision Tree and Gradient Boosted Tree is used for the second level of prediction[18].

The problems and obstacles that must be overcome while developing a system to identify credit card fraud are detailed in this paper. Additionally, feasibility analyses of several current fraud detection systems have been conducted and case studies addressing these systems have been conducted. The paper concludes with a detailed outline of future research objectives and opportunities for advancements in the field of credit/debit card fraud detection[19].

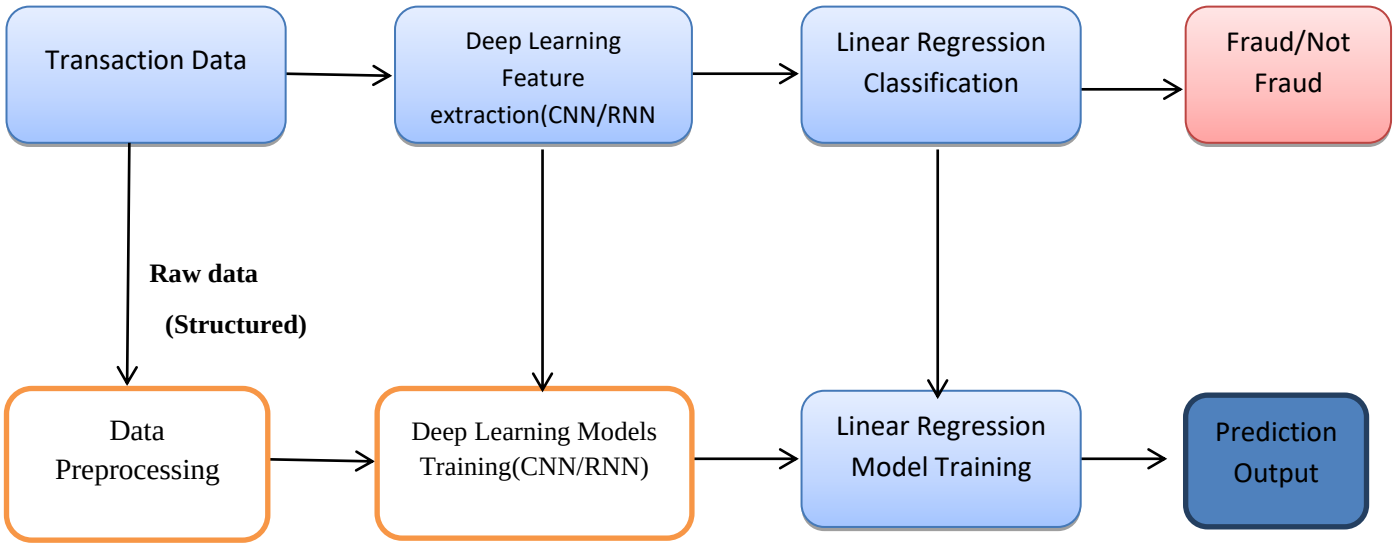

Fig 1 Block Diagram of the Overall System

Imbalanced Dataset and Incomplete Data

Data imbalance is one of the main things to consider when trying to identify credit card fraud. If one class in the data has a disproportionately high number of instances compared to the other classes, or if there is a significant disparity in the number of occurrences in each class, we say that the data is unbalanced.

Real data, not fake data, is what the fraud detection apps need to provide reliable findings. Due to confidentiality laws, dataset suppliers are unable to offer data directly. Data is anonymized before it is made public. In addition, the data will inevitably include inconsistent information, missing or NULL values, and unlawful values due to its real-time demand. The data will be misclassified if used in this way[1][4]. Therefore, it is important to sanitise the data before using it. Extraction, transformation, and loading are the three main stages. However, in practice, data use should dictate how the cleaning procedure is carried out[5].

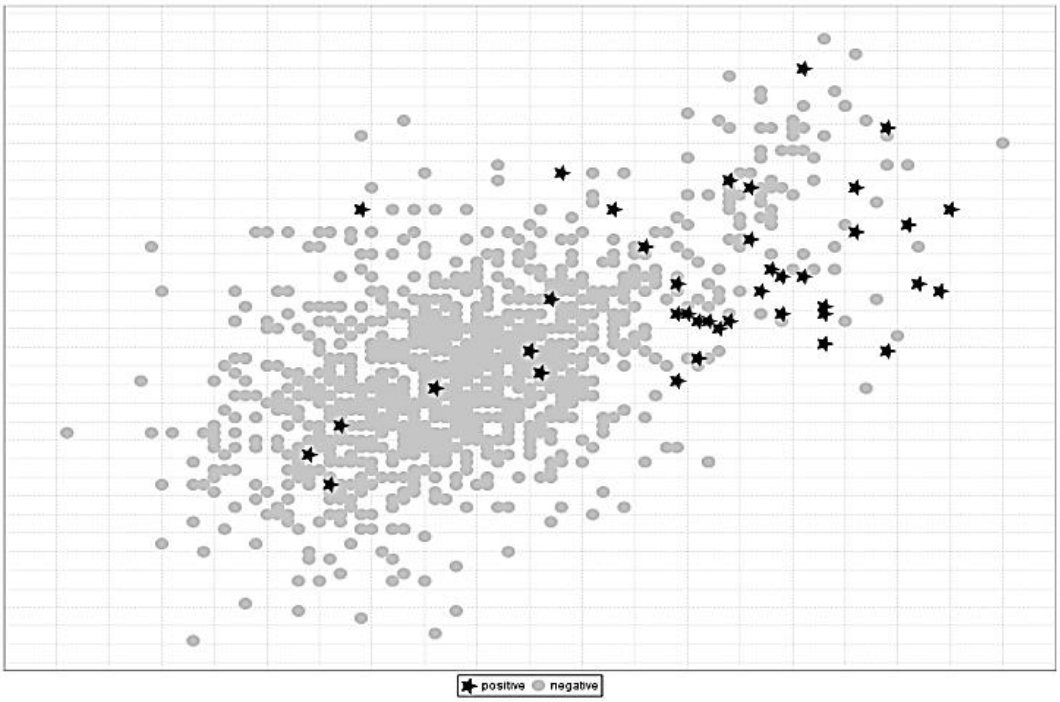

Fig 2. Data Imbalance – A Graphical View

Transaction Diversity

The consumer is yet another obstacle in the way of fraud detection. It is common practice to categorise fraud detection techniques as either "misuse based" or "anomaly based" depending on whether they identify patterns of fraud or not. The problem is that there are negative aspects to each of these approaches. While models based on abuse fail to detect new types of fraud, models based on anomalies fail to detect even small deviations from the pattern of a typical transaction. Even a seemingly ordinary transaction might be an abuse-based scam[8][9]. As far as fraud detection systems go, this is a major drawback. It is critical to build the system with minimal false positives since consumers are involved. As a result, banks end up settling for less than ideal conditions, which may lose them a handful of real fraud cases[11].

Real Scenario : Big Data

The availability of massive amounts of data and the rapidity with which it is being processed—together known as Big Data—make it difficult to design and implement a fraud detection system, even when using real-time data. The number of transactions has grown substantially as a result of the widespread use of electronic money. The amount of records received every unit of time has likewise increased significantly[12]. The detecting mechanism becomes more complicated as a consequence of this. In light of this, it is essential that any fraud detection system designed for the present day can swiftly handle massive amounts of information while yet producing reliable findings[14][15].

Emerging New Patterns of Fraud

Not only has widespread usage of technology increased throughout the last decade's technological boom, but so has its abuse. The system maintains equilibrium by executing fraudulent actions using more complex methods in response to the emergence of more sophisticated methods for detecting and preventing frauds. As new methods

and technology have emerged, disrupting this balance has become more difficult. While data mining and statistics are necessary for the first detection systems, machine learning and heuristics are required for the present situation. The greatest obstacle is the need for immediate solutions brought on by the problem's real-time nature[16].

Ensemble based Anomaly Detection in Credit/ Debit Card Transactions

Due to the significant degree of inconsistency in credit/debit card transactions, fraud detection has become an inherent part of these types of transactions. On the other hand, the domain does not mandate the ongoing verification of the legitimacy or fraud of a transaction. Since this can indicate a shift in the user's usual habits, authentication processes are performed before a transaction is marked as fraudulent. Because of this, the prediction model is essential for spotting outliers and setting off authentication processes. After authentication methods fail, frauds are permanently flagged[17]. This paper introduces an architecture for quickly and reliably detecting abnormalities in credit card transaction data via the use of an ensemble stacking method. Despite the fact that the regular classification issue model does not match to the standard classification method, anomaly detection in credit/debit card transactions is possible. All classes are given equal weight in a classification issue since the problem sees outcomes in terms of right and wrong predictions. When it comes to credit and debit card transactions, however, the importance of correctly anticipating a fraudulent transaction (the "true class") and a valid one (the "false class") is different. anticipating a scam with more accuracy is more important than anticipating the opposite. forecasting a valid transaction as fraudulent (False Negative) is much more dangerous than forecasting a fraudulent transaction as legal (True Positive). Because of this, a different operating mechanism than ordinary classifiers is required. Therefore, the best method for detecting anomalies in credit/debit card transactions is a stacking technique that integrates the classifiers' prediction skills with the option to include a bespoke final prediction mechanism, often known as a combiner algorithm[18].

Data pre-processing, feature aggregation, and ensemble rule development are the three stages that make up the suggested anomaly detection architecture for credit/debit card transactions. Data preparation for ensembles begins with pre-processing, and successive steps train and improve ensemble components, as well as construct the combiner algorithm for anomaly prediction. The following are the stages for the suggested model's algorithm and an explanation of it.[19]

1.Data Pre-processing

2. Feature Aggregation to form training data

3. Initialization and training of heterogeneous models

4. Prediction analysis based on predictability of fraud and legitimate transactions

5. Top k model selection based on the predictability of the model

6. Negative predictions from top k legitimate selection models is considered as the first phase of final predictions

7. Positive predictions from top k fraud selection models is considered as the second phase of final predictions

8. Discrepancy identification between positive and negative selection models

9. Rule based discrepancy correction and final prediction identification

10. Prediction aggregation to provide final predictions

Advantages of the Deep Learning

- **Enhanced Detection Accuracy**: By leveraging deep learning for feature extraction, the model can capture complex and non-linear patterns that traditional machine learning models might miss.
- **Improved Interpretability**: The use of linear regression for the final classification ensures that the model's predictions are interpretable, providing insights into the factors contributing to the detection of fraudulent transactions.

- **Robustness and Adaptability**: The hybrid approach can adapt to evolving fraud patterns and maintain high performance over time.

Conclusion

By integrating the best features of deep learning with linear regression, this research offers a better approach to detecting credit card fraud. Improved detection accuracy, interpretability, and resilience against changing fraud trends are all benefits of the suggested hybrid strategy, which overcomes the drawbacks of conventional machine learning models. An effective and practical solution for real-world applications in the financial sector is provided by the suggested technique, which leverages the sophisticated capabilities of deep learning for feature extraction and the simplicity of linear regression for classification. This method is a huge improvement above previous attempts at preventing credit card theft and keeping financial transactions secure

References

[1]Abdallah, A., Maarof, M.A. and Zainal, A., 2016. Fraud detection system: A survey. Journal of Network and Computer Applications, 68, pp.90-113.

[2] Akila, S. and Reddy, U.S., 2018. Cost-sensitive Risk Induced Bayesian Inference Bagging (RIBIB) for credit card fraud detection. Journal of computational science, 27, pp.247-254.

[3] Akoglu, L., Tong, H. and Koutra, D., 2015. Graph based anomaly detection and description: a survey. Data mining and knowledge discovery, 29(3), pp.626-688.

[4] Aleskerov, E., Freisleben, B. and Rao, B., 1997, March. Cardwatch: A neural network based database mining system for credit card fraud detection. In Proceedings of the IEEE/IAFE 1997 computational intelligence for financial engineering (CIFEr) (pp. 220-226). IEEE.

[5] Allan, T. and Zhan, J., 2010, May. Towards fraud detection methodologies. In 2010 5th International Conference on Future Information Technology (pp. 1-6). IEEE.

[6] Alowais, M.I. and Soon, L.K., 2012, June. Credit card fraud detection: Personalized or aggregated model. In 2012 Third FTRA International Conference on Mobile, Ubiquitous, and Intelligent Computing (pp. 114-119). IEEE.

[7] Bahnsen, A.C., Aouada, D., Stojanovic, A. and Ottersten, B., 2016. Feature engineering strategies for credit card fraud detection. Expert Systems with Applications, 51, pp.134-142.

[8] Bahnsen, A.C., Stojanovic, A., Aouada, D. and Ottersten, B., 2013, December. Cost sensitive credit card fraud detection using Bayes minimum risk. In 2013 12th international conference on machine learning and applications (Vol. 1, pp. 333-338). IEEE.

[9] Bekirev, A.S., Klimov, V.V., Kuzin, M.V. and Shchukin, B.A., 2015. Payment card fraud detection using neural network committee and clustering. Optical Memory and Neural Networks, 24(3), pp.193-200.

[10] Bhatla, T.P., Prabhu, V. and Dua, A., 2003. Understanding credit card frauds. Cards business review, 1(6), pp.1-15.

[11] Bhattacharyya, S., Jha, S., Tharakunnel, K. and Westland, J.C., 2011. Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), pp.602-613.

[12] Bolton, R.J. and Hand, D.J., 2001. Unsupervised profiling methods for fraud detection. Credit scoring and credit control VII, pp.235-255.

[13] Bolton, R.J. and Hand, D.J., 2002. Statistical fraud detection: A review. Statistical science, pp.235-249.

[14] Breiman, L., 1996. Bagging predictors. Machine learning, 24(2), pp.123-140.

[15] Breiman, L., 2001. Random forests. Machine learning, 45(1), pp.5-32.

[16] Cao, L., Yang, D., Wang, Q., Yu, Y., Wang, J. and Rundensteiner, E.A., 2014, March. Scalable distance-based outlier detection over high-volume data streams. In 2014 IEEE 30th International Conference on Data Engineering (pp. 76-87). IEEE.

[17] Carneiro, N., Figueira, G. and Costa, M., 2017. A data mining based system for credit-card fraud detection in e-tail. Decision Support Systems, 95, pp.91-101.

[18] Chan, P.K., Fan, W., Prodromidis, A.L. and Stolfo, S.J., 1999. Distributed data mining in credit card fraud detection. IEEE Intelligent Systems and Their Applications, 14(6), pp.67-74.