Journal of Theoretical and Applied Information Technology
31st July 2024. Vol.102. No. 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

RELATED FEATURE SUBSET MODEL FOR CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING METHODS

KALPANA PAWASE[1], VAIBHAV VASUDEVARAO GIJARE[2], NAGAMANI CHIPPADA[3]
JAYAKRISHNA AMATHI[4] ,SANJAY CHABILDAS PATIL[5]

[1,2] MIT Academy of Engineering, Alandi (D),Pune.
[3]CSE Department, Koneru Laksmaiah Education Foundation(KLEF), Vaddeswaram, Guntur, Andhra Pradesh-522302.
[4]Computer Science and Business Systems, R.V.R.&J.C. College of Engineering, Guntur, Andhra Pradesh.
[5]Thakur College of Engineering Technology, Kandivali (E.), Mumbai.

E-mail: [1]aswadhati.sirisha@gmail.com [3]drnagamanicse@gmail.com [4]amathijayakrishna@gmail.com

ABSTRACT

The Internet has experienced exponential growth over the past decade. Subsequently, the prevalence and prominence of services such as e-commerce, swipe and pay, and online bill payment have increased. Subsequently, criminals have intensified their endeavors to compromise credit card transactions. In the event that consumers are billed for items they did not purchase, it is imperative for credit card companies to possess the capability to identify fraudulent transactions. Data Science and Machine Learning are indispensable for resolving problems of this nature; their significance cannot be overstated. An increasing number of customers are demanding more amenities from businesses. An instance of such convenience is the capability of conducting online product purchases. The objective of this study is to illustrate the application of machine learning in the construction of a credit card fraud detection dataset. The credit card fraud detection problem involves the incorporation of data from successful credit card transactions into models of previous transactions. It is possible to ascertain the legitimacy of a new transaction by employing these methods. The prevalence of credit card fraud has increased in tandem with advancements in electronic payment systems and e-commerce. Procedures for detecting credit card fraud must therefore be implemented. When employing machine learning techniques for credit card fraud detection, it is vital to exercise extreme caution when selecting the characteristics of fraudulent transactions. This research presents a Related Feature Subset Model for Credit card Fraud Detection (RFSM-CFD) for accurate detection of credit card frauds. Feature selection for the machine learning -based credit card fraud detection system is proposed in this research. This research achieves 98.8% accuracy in feature subset generation and 98.5% accuracy in credit card fraud detection. When compared to state-of-the-art models, the proposed fraud detection model demonstrates superior accuracy. The outcomes indicate that the proposed model outperforms conventional models.

Keywords: *Credit Card, Fraud Detection, Machine Learning, Feature Set, Subset Model, Transaction Details, Credentials, Attackers.*

1. INTRODUCTION

Theft and fraud using payment cards, like credit cards or debit cards continue to be serious problems. Many fraud detection techniques are deployed to fight this problem. Fraudsters can attack in a number of ways, including the actual card being stolen or lost, or the card being compromised while appearing to have made a valid purchase [1]. According to the 2020 Global Payments Report, credit cards surpassed e-wallets and bank transfers as the most popular form of electronic payment worldwide [2]. Credit card fraud has been on the rise alongside the popularity of using these cards. Credit card fraud is on the rise, and that has serious consequences for the banking

Journal of Theoretical and Applied Information Technology
31st July 2024. Vol.102. No. 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

sector [3]. In 2020, fraudulent charges on credit cards cost consumers throughout the world a whopping USD 29.76billion [4]. The credit card predictive model for fraud detection is shown in Figure 1.

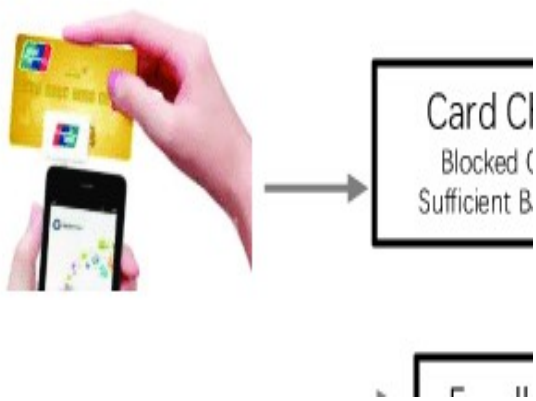

Fig 1: Credit Card Fraud Predictive Model

These days, banks and other financial organizations often create fraud detection systems that are tailored specifically to their own portfolios [5]. There has been a widespread use of machine learning and data mining methods for analyzing normal and abnormal behaviour patterns and individual transactions to identify potential instances of fraud [6]. The most efficient and low-cost strategy, given the current state of affairs, is to employ statistical algorithms to sift through accessible data in search of potential proof of fraud. In order to calculate a fraud probability estimattion to each transaction [7], autonomous models trained on labelled data analyze all labelled transactions in the past. The neural networks is one of the most common supervised methods, however alternative models such as support-vector-machines (SVMs) and decision trees have also been utilized [8].

Since a decade ago, the Internet's prominence has skyrocketed. As a consequence, the prevalence of e-commerce, tap-and-pay systems, online bill payment systems, and similar technologies has skyrocketed. Credit card fraud attacks have increased as a consequence [9]. Tokenization and encryption of credit card data are merely two of the numerous techniques used to protect credit card-related financial transactions. Although these precautions aid in the prevention of credit card fraud in a number of circumstances, they are not absolute. Machine learning (ML), a subfield of artificial intelligence (AI) [10], empowers machines to improve their predictive capabilities through experiential learning, without requiring explicit programming. Utilizing ML techniques to accurately detect credit card fraud and identify fraudulent transactions on credit cards is the objective of this research [11].

Considered credit card fraud occurs when an unauthorized individual uses a credit card to make a purchase. According to the Federal Trade Commission (FTC) [12], a total of 2119 data breaches compromised 243 million records, with credit card fraud constituting the predominant type of compromise. Thus, it is critical to have a dependable method for detecting credit card fraud that can safeguard users' funds from theft [13]. The fact that the majority of models applying ML techniques to the problem of credit card fraud detection cannot be replicated is a significant obstacle. This is required due to the private nature of credit card usage [14]. Certain details regarding the underlying data are thus concealed in the datasets utilized to train ML models for credit card fraud detection. Moreover, the ever-changing nature of fraudulent transactions, including their structure and patterns, complicates the detection of credit card fraud [15]. An additional challenge that remains unresolved by existing machine learning models utilizing illicit credit card transactions is the dataset's highly skewed distribution [16]. In order to achieve this objective, it is vital to develop ML models that operate effectively and can detect instances of credit card fraud with precision. Some of the supervised machine learning algorithms used to detect credit card fraud [17] include Decision Tree (DT), Random Forest (RF), Artificial Neural Network (ANN), Naive Bayes (NB), and Logistic Regression.

Large datasets are used for training and testing ML systems. The research makes use of a dataset of fraudulent credit card transactions culled from the accounts of cardholders available in public data set providers [18]. There is a risk that the training quality of the classifiers will suffer if the dataset has a large number of attributes. Because of the widespread availability of the internet now, digital statistics are easily accessible everywhere in the world. Cloud storage allows businesses of all sizes to store their data, regardless of how large or little they are. Social media followers, order patterns [19], likes, and shares are just a few examples of the abundant data sources available today. The financial sector, commercial institutions, and governments are all increasingly vulnerable to the effects of white-collar crime, which is on the rise.

The practice of deceiving another person for financial gain is known as fraud. The advantages of combining communication technology with improved card transactions are starting to sink in for more and more people [20]. Unfortunately, card

fraud is becoming more common as credit card payments become the standard for both in-person and online transactions. Because it eliminates the need for humans to physically handle huge datasets, machine learning is a technological marvel of the 21st century [21]. There have been a lot of monitored ways used to find credit card theft in the past few decades. One major challenge in using ML for fraud detection is dealing with extremely imbalanced datasets [22]. Existing evidence sets demonstrate that fraud is incredibly unlikely due to the preponderance of valid transactions. Researchers have significant challenges in developing an efficient and effective detection and prevention framework that has a low false positive rate and a high detection rate [23].

A machine learning-based method for detecting fraudulent transactions on credit cards and providing issuers with feedback is proposed in this work. This feedback strategy enhances both the detection accuracy and overall performance of the classifier. Conduct an evaluation of the effectiveness of various classification strategies, including random forest, tree classifiers, artificial neural networks, support vector machine [24], Naive Bayes, logistic regression, and gradient boosting classifier approaches, on a credit card fraud database that is significantly skewed. Organizations and banking institutions commonly use machine learning for fraud detection because of its effectiveness in spotting suspicious financial activities. For a number of reasons, though, machine learning may struggle with fraud detection. As the amount of fraudulent charges is extremely low [25], the data distribution is highly skewed, the data is dynamic and ever-changing, and there is a dearth of real-world datasets due to privacy issues. Multiple strategies were presented to address these difficulties, with a primary emphasis on adopting a hybrid paradigm [26]. Several reports have discussed the development of optimization algorithms for fraud detection [27]. To confirm that the specified model is the best option for the specified dataset, these hybrid models solely used a model without considering the performances of other models.

Incorrectly labeling valid transactions as fraudulent leads to false positives. When valid transactions are either refused or need further verification due to a high false positive rate, it can cause consumer displeasure, inconvenience, and even business loss.It is of the utmost importance to guarantee data security and adhere to legislation like GDPR. The financial and legal ramifications of a data breach can be devastating. A lack of interpretability makes it difficult to comprehend the rationale behind a fraudulent transaction flag, which in turn reduces confidence in the system and makes it more difficult to comply with rules that demand justifications for actions.

To overcome the limitations, the primary contribution of this research is the creation and examination of numerous models with the same dataset, with the goal of identifying a ML based model on the assessment of its performance prediction. Because most fraud datasets, and credit card datasets in particular, are labelled, this research is limited to the use of categorization supervised machine learning for the purpose of detecting credit card fraud. This research presents a Related Feature Subset Model for Credit card Fault Detection (RFSM-CFD) for accurate detection of credit card frauds.

2. LITERATURE SURVEY

Customers find credit card processing for online purchases to be a time-efficient and practical alternative. The prevalence of plastic has increased in tandem with the risk of credit card fraud. Credit card fraud causes financial institutions and victims to incur substantial losses. This work has been primarily concerned with detecting such frauds due to factors including readily accessible public data, high-class imbalanced data, the changing nature of fraud, and high false alarm rates. The literature mentioned above contains descriptions of numerous machine learning-based methods utilized for credit card identification. These methods include, but are not limited to, the Xtreme Learning Approach, Decision-Tree, Random-Forest, Support-Vector Machine, Logistic Regression, and XG Boost. Nevertheless, state-of-the-art deep learning algorithms continue to be necessary in order to enhance precision and reduce fraudulent losses. Using the recently created deep learning algorithms that were made specifically for this purpose has been the main goal. After comparing deep learning and machine learning methods, Alarfaj et al. [1] were able to achieve successful results. The European Card Standard dataset is utilized in a comprehensive empirical analysis aimed at detecting fraud. The dataset was initially processed using a machine learning method, which improved the reliability of the fraud detection procedure. To further enhance the capacity to detect fraud, three topologies built on convolution neural networks are then employed. Further improvement in detection accuracy was achieved by adding more layers. To carry out a thorough empirical investigation,

experiments were carried out with different numbers of hidden layers, time periods, and the most current models.

Credit cards are now more often used for everyday expenditures, thanks to developments in communication and online shopping. Nevertheless, there has been a noticeable uptick in credit card theft, which results in significant yearly losses for financial institutions. It is difficult to create fraud prevention algorithms that adequately reduce these costs since most credit and debit card datasets are extremely biased. Additionally, traditional ML systems' dependence on a static translation between the input and output vectors leads to subpar performance when it comes to identifying credit card fraud. Consequently, their levels of adaptability aren't enough to deal with the ever-changing buying habits of credit card users. One solution that has been suggested by Esenogho et al. [2] to effectively detect credit card fraud is to combine a computational model ensemble classifier with a hybrid data oversampling technique. An LSTM neural network serves as the foundational learner in the ensemble classifier that is generated using the AdaBoost algorithm. The SMOTE-ENN approach and the Synthesis Minority Oversampling Techniques are used for hybrid resampling, on the other hand.

With the rise of e-commerce and other forms of FinTech (financial technology), the use of credit cards for online purchases has become more commonplace. Credit card fraud has thus been on the rise, impacting businesses, financial institutions, and the companies that supply the cards. Consequently, models for the security of credit card transactions that successfully prevent theft and fraud are of the utmost importance. In order to create an ML system for fraud detection, Ileberi et al. [6] used real-world unbalanced datasets collected from European credit cardholders. The class imbalance was fixed by resampling the dataset using the Synthesis Minority Oversampling Technique (SMOTE).Support Vector Machines (SVMs), Logistic Regression (LRs), Random Forests (RFs), Gradient Boosting (XGBoost), Decision-Tree (DT), and Extra-Tree (ET) were among the machine learning methods used to assess the system's performance. Applying the Adaptation Boosting (AdaBoost) method improved the classification accuracy of the ML systems. Models were assessed using AUC, Accuracy, Precision, Recall, and Matthews Correlation Coefficients (MCC).

Credit cards have been widely accepted as a payment option for both in-store and online transactions due to the expansion of modern e-commerce systems and communications technologies. Unfortunately, this has also led to a dramatic increase in credit card fraud. Every year, businesses and consumers suffer enormous losses as a result of credit card theft. Identity thieves are always looking for new ways to steal sensitive information. More widespread use of electronic payment systems is being slowed down by the urgent need to detect fraudulent transactions. Credit card fraud can only be detected with methods that are both quick and accurate. This research by Benchaji et al. [8] improved a light gradients boosting machine (OLightGBM) to identify fraudulent credit card transactions. Smartly using a Bayesian-based hyperparameter optimization method, the suggested method optimizes the configuration of a lightweight gradient boosting machine (LGBM). The author conducted trials using two publicly available credit card transaction datasets to verify that the suggested OLightGBM could successfully identify fraudulent transactions.The categorization of information that is unbalanced by class has garnered increasing interest from scientists in numerous scientific disciplines, such as metabolic engineering, cancer diagnosis, and fraud detection, among others, in recent years. In order to enhance classification outcomes, Fatima et al. [9] proposed a technique that selects features with minimal data overlap by utilizing an enhanced R-value.The author presented three feature selection methods, namely Reduced Overlapping with ADASYN (ROA), Reduced Overlapping with SMOTE (RONS), and Reduced Overlapping with No-samplings (ROS), all of which are based on dense feature selection in order to minimize overlap and perform binary classification. An additional parallel between ROS and ROA pertains to the incorporation of a resampling procedure. As feature selection approaches, the simulation results indicate that the variation in the false discovery rate during the main feature selection for process modeling is effectively managed. Furthermore, this effective feature selection method can be extended as a substitute approach to address the challenge of overlapping class imbalances in learning.

Credit and debit card fraud is a significant issue that can deplete funds from the accounts of cardholders and severely financially impact card issuers. In order to detect fraudulent activities in financial records, contemporary methods employ machine learning algorithms. As the manual generation of features necessitates domain expertise and potentially serves as the impetus for fraudulent

strategies, it is imperative that the online detection system autonomously identifies the most significant patterns of fraudulent activity. In order to detect fraudulent credit card transactions, Cheng et al. [10] proposed spatial-temporal attention-based graph networks (STAGN) in this study. More precisely, a graph learning algorithm is employed to initially acquire knowledge of the temporal and spatial attributes of the transaction graph. Following this, the learned tensor representations are fed into a network of convolution layers, and the output is subjected to spatial-temporal attention. In conjunction with detection and end-to-end 3D convolutional networks, attentional weights are trained. Extensive investigations are conducted using the complete dataset comprising actual card transactions. STAGN exhibits superior performance compared to other contemporary boundaries, as evidenced by its precision and recall curves as well as AUC. Furthermore, the author conducted empirical investigations in collaboration with subject matter experts regarding the proposed protocol for knowledge discovery and fraud detection. The findings indicate that the protocol outperforms other approaches in terms of identifying fraudulent patterns, mining temporal and spatial fraud hotspots, and detecting fraudulent transactions. The effectiveness of the proposed strategy is demonstrated in the context of additional user behavior pattern activities. In an ultimate endeavor to address the challenges posed by big data, the author detailed the architecture of each system module and incorporated the suggested STAGN into the fraud detection algorithm as the predictive model.

Significant academic efforts have been put into discovering new ways to spot various forms of fraud. These approaches, however, have been shown to be futile. Due to the low proportion of fraudulent transactions relative to legitimate ones, spotting them with a classification algorithm is challenging in this scenario. Makki et al. [14] conducted a comprehensive experimental study of the proposed solutions to the unbalancing classification problem in this paper. Both these options and the computer learning algorithms currently in use for fraud detection were investigated. The author used a fraudulent activity labelled dataset to determine their limitations and summarize the findings. This paper argues that imbalanced classification approaches fail to perform well, particularly when the data are highly unbalanced.

3. PROPOSED MODEL

Credit card fraud is a serious ethical problem in the business world. The research primary goal is to detect credit card fraud and offer a workable solution to this problem. Credit card fraud has resulted in billions of dollars in losses worldwide for victims and financial institutions [28]. Even though there are many security measures in place to prevent fraud, attackers are always devising new techniques to deceive unsuspecting victims. The banking industry and other financial institutions place a premium on fraud detection [29]. The proposed modeluse historical fraud data to improve its ability to identify future instances of fraud. While fraud detection algorithms based on mining data had been tried, they had not shown any promising results. The research makes use of supervised learning methods applied to a highly skewed and imbalanced dataset considered from Kaggle.

The process of feature selection (FS) is essential in the application of machine learning techniques. This is in part due to the fact that the datasets used for training and testing may have a huge feature space, which might have a detrimental effect on the models' overall performance. The nature of the research challenge will dictate the FS technique of choice [30]. Artificial intelligence, namely pattern matching, is used in several methods for detecting fraud. It is crucial that fraud be uncovered through safe and effective means.The range, unit, and level of most features in a real-world dataset will be different. When the level of one aspect is far larger than the others, it tends to overshadow them all. So, to make use of classification methods and get rid of the effect of different quantitative units, raw data should be scaled. Because of this, the MinMaxScaler method was implemented to rescale the features between 0 and 1 for this research. When fraud cannot be avoided, however, it must be uncovered quickly so that appropriate measures can be taken. Identifying whether or not a transaction is authentic is the goal of fraud detection. Since it would be impossible for humans to manually check each and every transaction to determine if it is fraudulent or not, automated fraud detection solutions are essential. The credit card fraud detection general process is shown in Figure 2.

Journal of Theoretical and Applied Information Technology
31st July 2024. Vol.102. No. 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

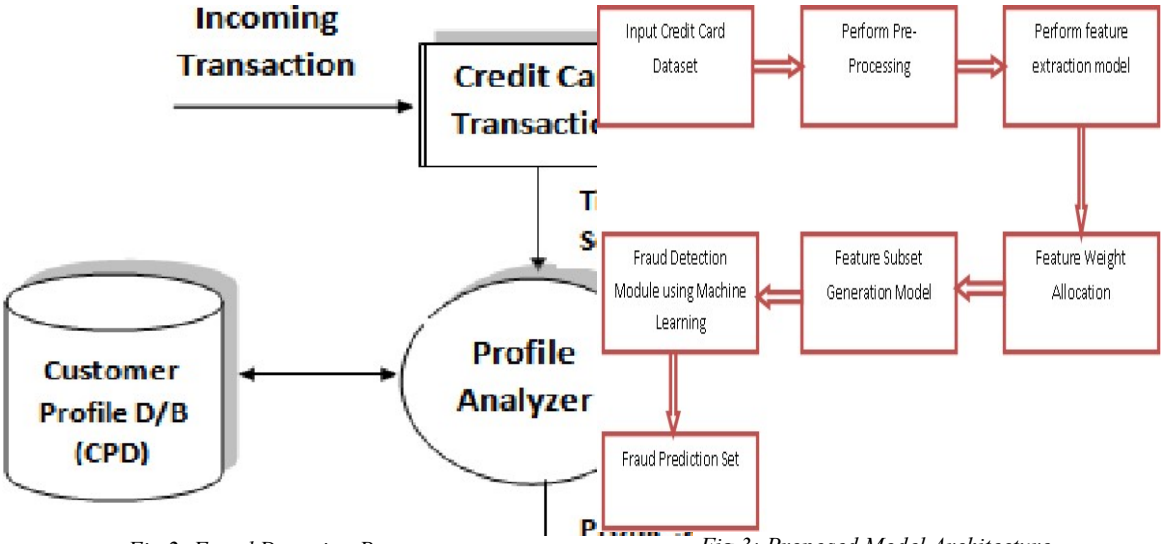

Fig 2: Fraud Detection Process

Fig 3: Proposed Model Architecture

Feature selection is a crucial part of data preparation to avoid over fitting due to the curse of dimensionality reduction. Through the process of feature selection, superfluous or unimportant details are eliminated. Filter and wrapper are the two most well-known approaches, and both have their advantages and drawbacks. The wrapper method has some drawbacks, such as the high processing cost and reliance on the algorithm as an evaluation function to select the features. On the other hand, the filter approach has the drawback of only searching for features independently, therefore features that are highly dependent on one another will be missed.

A ML based feature selection strategy that combines filter and wrapper approaches is an alternate solution that is less exhaustive and has less drawbacks. In order to determine the degree of similarity between the numerical characteristics, a correlation-based filter was employed. Positive features that were highly associated were omitted from the prediction model to prevent over fitting and to conserve computational resources. This research presents a Related Feature Subset Model for Credit card Fault Detection for accurate detection of credit card frauds. The proposed model architecture is shown in Figure 3.

Algorithm RFSM-CFD

{

Input: Credit Card Fraud Dataset {CCF_{DSET}}

Output: Fraud Prediction Set {FP_{SET}}

Step-1: Load the dataset and then analyze the records to perform pre processing on the dataset. The pre processing cleans the data from the available ones. The pre processing is performed as Consider records in dataset as {R_1, R_2,……,R_N} and the instance of data record fields as {I_1, I_2,……, I_N}.

A $\log_{10} R_I$ is used if the data response is linear to the function $CCF_{DSET}(X)$. Because the log of negative numbers results in undefined values like Not a Number (NaN), this procedure first sets negative numeric value to zero. This impact can be avoided by doing an integrand preprocessing step before performing a Log10 step.

$$DPreSet = \int_{R=1}^{M} \frac{getval(R)}{\log_{10} R_I}$$

$$DPreFSet = \log_{10}\left(\frac{1}{\max(DPreSet)}\right)$$

$$Mean_{Set} = \sum_{R=1}^{M} mean\left(\frac{I_R}{I_{R+1}}\right)$$

Step-2: As a method of dimensionality reduction, feature extraction organises large amounts of raw data into more manageable parts. In order to process these massive data sets, a great deal of computational power is needed because of the sheer amount of variables involved. To reduce the amount of data that needs to be processed while still providing an accurate and complete description of the original data set, a number of techniques

Journal of Theoretical and Applied Information Technology
31st July 2024. Vol.102. No. 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

have been developed under the feature extraction process. The feature extraction process is performed as

$$FEset(R(i)) = \sum_{R=1} \frac{CCFDset(R) + mean(R(I_R), CCFDset(R+1)}{\max(CCFDset(R))} + \sum_{R=1}^{N} \frac{getattr(\max(Mean(R), Mean(R+1))}{len(CCFDset)}$$

Step-3: The parameters employed in each layer of the model are reflected in the model's weights. The weights are allocated based on the correlation factor and the highly correlated features are removed and most useful features are considered. The feature weight allocation process is performed as

$$Fweight(CCFDset[M]) = \sum_{R=1}^{M} \frac{\max\left(FEset(R(i))\right) + \min(FEset(R, R+1)) - \sum_{R=1}^{M} \max(\text{sim}(R_{(R)}, R_{(R+1)})}{\text{len(FEset)}} + \sum_{R=1}^{M} \frac{\max(R(R-1)) - \min(R(R))}{\text{size}(Mean_{Set})}$$

Step-4: From the features extracted, the feature subset is generated which considers the most useful features that is used for credit card fraud detection. The feature subset generation process is performed as

$$Fsubset(R[M]) = \frac{\sum_{R=1}^{M} \max(corr(FEset(R), FEset(R-1))}{\sum_{R=1}^{M} \sqrt{\frac{\max(FWeight(R+1)) - \min(FWeight(R)))}{TotalTim \min(D\,\text{Pr}\, eFset(R))}}} + \max(Meanset(R))$$

Here corr is the correlation function used to calculate the correlated featuers.

Step-5: The credit card fraud detection is performed by training the model using the feature subset and the final fraud prediction set is generated as

$$FP[M] = \sum_{R=1}^{M} \frac{\max(Fsubset(R)) - \min(FWeigh(R)}{len(Fsubset(R))} + \frac{\min(corr(R+1, R)) - \max(FWeigh(R+1))}{M}$$

$$FPset[M] = \sum_{R=1}^{M} \frac{\max(Fsubset(R+1))}{len(Fsubset(R))} \begin{cases} if\ Fsubset \leftarrow \min(Corr(R)) & FP \leftarrow 0 \\ otherwise & FP \leftarrow 1 \end{cases}$$

}

4. RESULTS

Machine learning algorithms can identify suspicious activity on a credit card and prevent further losses. The first step in using a model to predict potential instances of fraud is to collect and organize raw data for use in training that model. Machine learning provides solutions for detecting credit card fraud, including the use of learning algorithms to classify transactions as authentic or fraudulent, credit card profiling to predict whether legitimate cardholders or malicious actors are using the cards, and outlier detection methods to identify records of transactions that are significantly different from the norm.

Credit card fraud is a big problem in today's interconnected global economy. Worldwide, fraud results in massive financial losses. As a result, many banks have invested in studying the problem and creating tools to help identify and stop credit card fraud. The major goal of this research is to develop a ML model that efficiently and accurately identify fraudulent transactions for credit card issuers. A computer software that may rapidly construct a prediction system for detecting credit card fraud by automatically picking suitable Machine Learning algorithms, adjusting their hyper-parameter variables, and evaluating performance on a highly skewed dataset.

There is zero overhead in terms of user setting of method parameters during model training. It also requires little work to apply the model, retrain this whenever new data becomes available, produce visualizations of the findings, and communicate them across the many levels of management in the organization. This research presents a Related Feature Subset Model for Credit Card Fault Detection (RFSM-CFD) for accurate detection of credit card frauds. Feature selection for the machine learning -based credit card fraud detection system is proposed in this research. The proposed model is implemented in python and executed in Google Colab. The proposed model is compared with the traditional Enhanced credit card fraud detection based on attention mechanism and LSTM deep model (FDAM-LSTM) model.

Data preprocessing, or the manipulation or removal of data before to its usage, is a crucial part of the data mining process, as it ensures or improves performance. Any action taken on raw data in order

Journal of Theoretical and Applied Information Technology
31st July 2024. Vol.102. No. 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

to get it ready for further processing is referred to as data preprocessing, and it is part of the larger data preparation process. It is a crucial first stage in the data mining process and has been for a long time. The Figure 4 shows the data pre processing accuracy levels of the existing and proposed models.

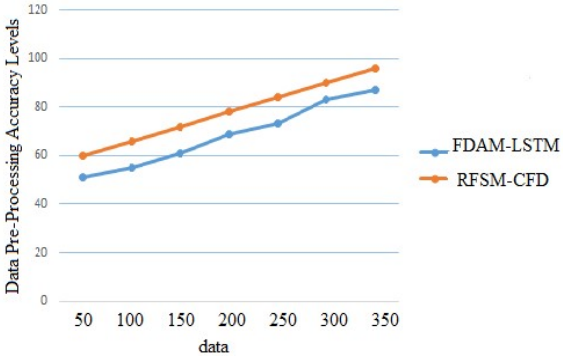

Fig 4: Data Pre-Processing Accuracy Levels

The process of extracting features to be used in machine learning for training the model. The process of transforming unstructured data into a set of quantifiable features that can be further processed without losing any of the context of the original data is referred to as feature extraction. The feature extraction is used to describe the operation. Using machine learning on data that has already been preprocessed does not provide results that are as excellent as using it on raw data. The features extracted will be used as input for learning models. The feature extraction time levels of the existing and proposed models are shown in Figure 5.

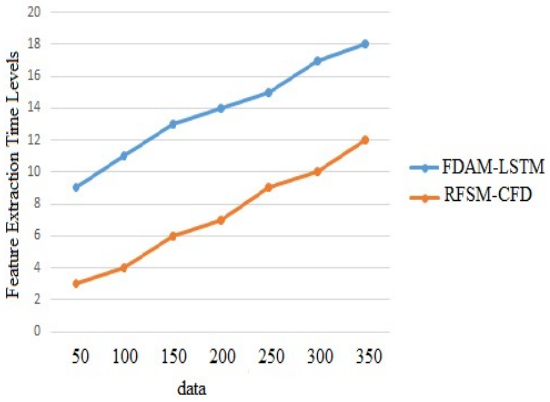

Fig 5: Feature Extraction Time Levels

On a scale ranging from one to thresholds, the feature weights are ranked. A feature weight of zero implies that the feature must be treated as available space, whereas a feature weight of maximum indicates that the feature is considered an impediment and should not have labels overlap it. The feature weight represents which feature to consider and which features to remove. The feature weight allocation accuracy levels of the traditional and proposed models are shown in Figure 6.

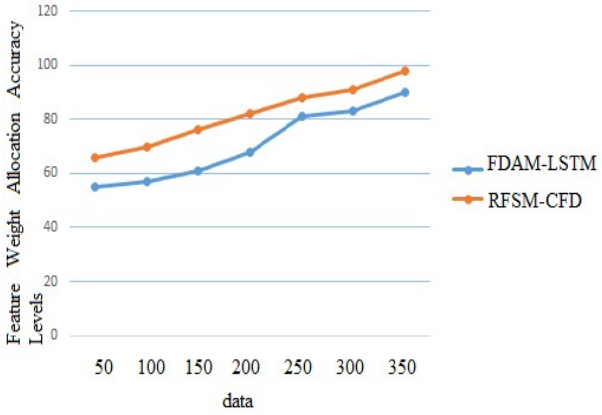

Fig 6: Feature Weight Allocation Accuracy Levels

In the preliminary processing phase of machine learning, feature selection is the most critical stage. The idea is to focus on the attributes or features that contribute the greatest value to a machine learning activity. A subset generation method is a search method that selects feature subsets using certain search strategies such as sequential search or random search. Using a the proposed approaches, we estimate a subset of characteristics depending on a variety of criteria. The Figure 7 ndicates the feature subset generation time levels of the proposed and the traditional method.

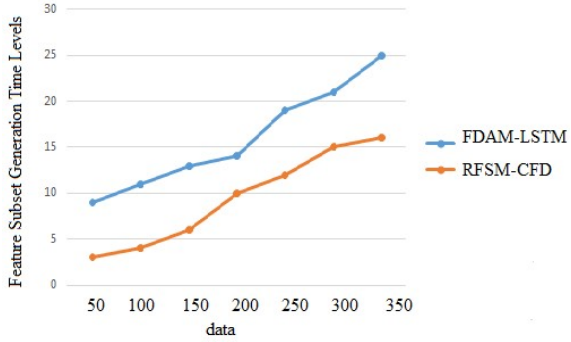

Fig 7: Feature Subset Generation Time Levels

Feature generation pertains to the procedure by which additional features are constructed from pre-existing ones, with the explicit purpose of incorporating them into the statistical analysis. Typically, this entails collecting supplementary data that enhances the predictive capabilities of the model. When two or more features interact, the generation of novel features may be performed to improve model precision. In machine learning and statistics, feature selection, alternatively referred to as attribute selection, variable subset selection, or

Journal of Theoretical and Applied Information Technology
31st July 2024. Vol.102. No. 14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

spatial selection, is a method employed to reduce an extensive set of prospective features to a more feasible subset comprising practical variables and predictors. The feature subset generation accuracy levels of the traditional and proposed models are shown in Figure 8.

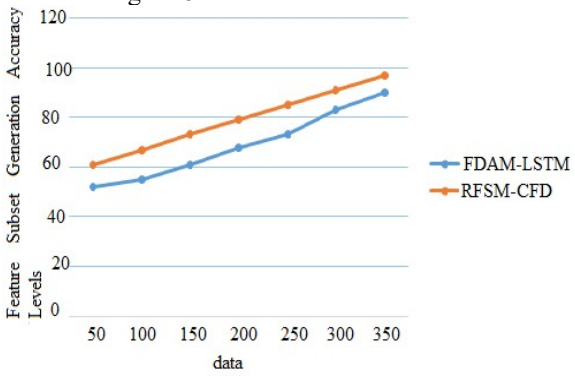

Fig 8: Feature Subset Generation Accuracy Levels

Detecting fraudulent use of a credit card involves determining when a purchase attempt is fraudulent rather than actually completing the transaction. Most businesses use a combination of fraud detection technologies and methods because no single method is 100% effective. The proposed model fraud detection time levels is less than the existing models that is represented in Figure 9.

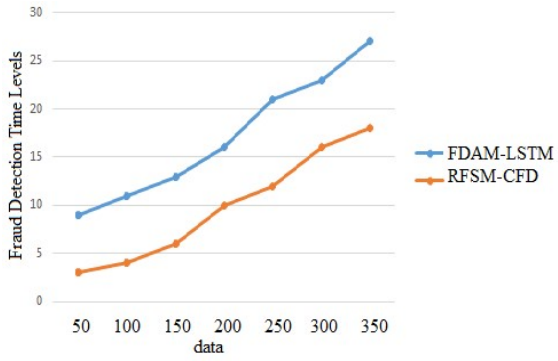

Fig 9: Fraud Detection Time Levels

Detecting fraudulent use of a credit card involves determining when a purchase attempt is fraudulent rather than actually completing the transaction. In the final analysis, we consider two key aspects of credit card fraud detection: the size of a transaction and the interval between related transactions. For spotting fraudulent activity, these characteristics are excellent indicators. The fraud detection accuracy levels of the proposed model is high than the existing models that is illustrated in the Figure 10.

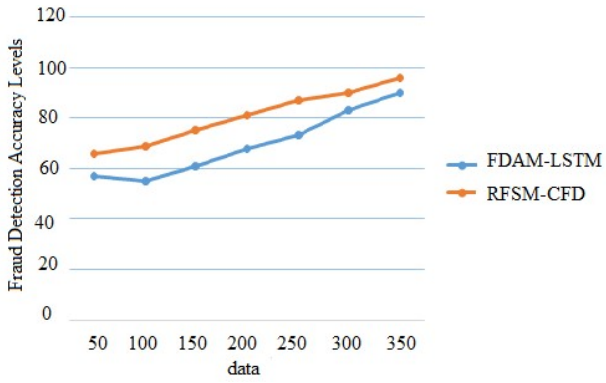

Fig 10: Fraud Detection Accuracy Levels

5. CONCLUSION

In this research, the feasibility of employing linear and nonlinear machine learning models on data collected from credit card transactions is analyzed. To ascertain which purchases are most likely to be fraudulent, the proposed model mainly control such fraud models. The obtained model can then be used by a fraud detection system. While sophisticated nonlinear algorithms are readily available, it has been observed that well-designed based approaches are formidable adversaries. The creation of high-quality expert characteristics that can fully capture the indicators of the problem's behavior into smart variables is a vital step. Card fraud, whether involving credit, debit, or both, has been on the rise in recent years. It has come to light that many credit card users have been duped into giving out their personal information, card numbers, and one-time passwords in response to unsolicited phone calls. People's birthdates, vehicle identification numbers, the year they met their mothers, the year they had their first real life experience, etc. are all examples of basic stick mystery keys used by a wide range of consumers. This sort of password is easily cracked using mining algorithms. In order to prevent this, an automatic anti-deception framework that can correctly identify a customer's approval is required. To avoid financial loss, businesses can use one of several anti-fraud methods or software. This research presents a Related Feature Subset Model for Credit card Fault Detection for accurate detection of credit card frauds. This research achieves 98.8% accuracy in feature subset generation and 98.5% accuracy in credit card fraud detection. In future, feature reduction strategies will also be employed at multi level fraud detection to reduce the time complexity and improve the system performance.

REFERENCES

[1]. F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan and M. Ahmed, "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms," in IEEE Access, vol. 10, pp. 39700-39715, 2022, doi: 10.1109/ACCESS.2022.3166891.

[2]. E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba and G. Obaido, "A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection," in IEEE Access, vol. 10, pp. 16400-16407, 2022, doi: 10.1109/ACCESS.2022.3148298.

[3]. A. Zhakov et al., "Application of ANN for Fault Detection in Overhead Transport Systems for Semiconductor Fab," in IEEE Transactions on Semiconductor Manufacturing, vol. 33, no. 3, pp. 337-345, Aug. 2020, doi: 10.1109/TSM.2020.2984326.

[4]. J. Al Hage, P. Xu, P. Bonnifait and J. Ibanez-Guzman, "Localization Integrity for Intelligent Vehicles Through Fault Detection and Position Error Characterization," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 4, pp. 2978-2990, April 2022, doi: 10.1109/TITS.2020.3027433.

[5]. E. Wu, H. Cui and R. E. Welsch, "Dual Autoencoders Generative Adversarial Network for Imbalanced Classification Problem," in IEEE Access, vol. 8, pp. 91265-91275, 2020, doi: 10.1109/ACCESS.2020.2994327.

[6]. E. Ileberi, Y. Sun and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost," in IEEE Access, vol. 9, pp. 165286-165294, 2021, doi: 10.1109/ACCESS.2021.3134330.

[7]. E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba and G. Obaido, "A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection," in IEEE Access, vol. 10, pp. 16400-16407, 2022, doi: 10.1109/ACCESS.2022.3148298.

[8]. Benchaji, Ibtissam & Douzi, Samira & Ouahidi, Bouabid & Jaafari, Jaafar. (2021). Enhanced credit card fraud detection based on attention mechanism and LSTM deep model. Journal of Big Data. 8. 10.1186/s40537-021-00541-8.

[9]. E. B. Fatima, B. Omar, E. M. Abdelmajid, F. Rustam, A. Mehmood and G. S. Choi, "Minimizing the Overlapping Degree to Improve Class-Imbalanced Learning Under Sparse Feature Selection: Application to Fraud Detection," in IEEE Access, vol. 9, pp. 28101-28110, 2021, doi: 10.1109/ACCESS.2021.3056285.

[10]. D. Cheng, X. Wang, Y. Zhang and L. Zhang, "Graph Neural Network for Fraud Detection via Spatial-Temporal Attention," in IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 8, pp. 3800-3813, 1 Aug. 2022, doi: 10.1109/TKDE.2020.3025588.

[11]. Z. Zhang and G. Yang, "Distributed Fault Detection and Isolation for Multiagent Systems: An Interval Observer Approach," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 50, no. 6, pp. 2220-2230, June 2020, doi: 10.1109/TSMC.2018.2811390.

[12]. Z. Yin, L. Wang, B. Zhang, L. Meng and Y. Zhang, "An Integrated DC Series Arc Fault Detection Method for Different Operating Conditions," in IEEE Transactions on Industrial Electronics, vol. 68, no. 12, pp. 12720-12729, Dec. 2021, doi: 10.1109/TIE.2020.3044787.

[13]. W. Jia and J. Wang, "Partial-Nodes-Based Distributed Fault Detection and Isolation for Second-Order Multiagent Systems With Exogenous Disturbances," in IEEE Transactions on Cybernetics, vol. 52, no. 4, pp. 2518-2530, April 2022, doi: 10.1109/TCYB.2020.3007655.

[14]. S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. Hacid and H. Zeineddine, "An Experimental Study With Imbalanced Classification Approaches for Credit Card Fraud Detection," in IEEE Access, vol. 7, pp. 93010-93022, 2019, doi: 10.1109/ACCESS.2019.2927266.

[15]. H. Tingfei, C. Guangquan and H. Kuihua, "Using Variational Auto Encoding in Credit Card Fraud Detection," in IEEE Access, vol. 8, pp. 149841-149853, 2020, doi: 10.1109/ACCESS.2020.3015600.

[16]. R. San Miguel Carrasco and M. -Á. Sicilia-Urbán, "Evaluation of Deep Neural Networks for Reduction of Credit Card Fraud Alerts," in IEEE Access, vol. 8, pp.

186421-186432, 2020, doi: 10.1109/ACCESS.2020.3026222.

[17]. S. N. Kalid, K. -H. Ng, G. -K. Tong and K. -C. Khor, "A Multiple Classifiers System for Anomaly Detection in Credit Card Data With Unbalanced and Overlapped Classes," in IEEE Access, vol. 8, pp. 28210-28221, 2020, doi: 10.1109/ACCESS.2020.2972009.

[18]. B. Lebichot, T. Verhelst, Y. -A. Le Borgne, L. He-Guelton, F. Oblé and G. Bontempi, "Transfer Learning Strategies for Credit Card Fraud Detection," in IEEE Access, vol. 9, pp. 114754-114766, 2021, doi: 10.1109/ACCESS.2021.3104472.

[19]. B. Can, A. G. Yavuz, E. M. Karsligil and M. A. Guvensan, "A Closer Look Into the Characteristics of Fraudulent Card Transactions," in IEEE Access, vol. 8, pp. 166095-166109, 2020, doi: 10.1109/ACCESS.2020.3022315.

[20]. Lakshman Narayana,V., Lakshmi Patibandla, R.S.M., Pavani, V., Radhika, P. (2023). Optimized Nature-Inspired Computing Algorithms for Lung Disorder Detection. In: Raza, K. (eds) Nature-Inspired Intelligent Computing Techniques in Bioinformatics. Studies in Computational Intelligence, vol 1066. Springer, Singapore. https://doi.org/10.1007/978-981-19-6379-7_6.

[21]. Narayana, V.L., Sujatha, V., Sri, K.S., Pavani, V., Prasanna, T.V.N., Ranganarayana, K. (2023). Computer tomography image based interconnected antecedence clustering model using deep convolution neural network for prediction of covid-19. Traitement du Signal, Vol. 40, No. 4, pp. 1689-1696. https://doi.org/10.18280/ts.400437

[22]. H. Zhou, G. Sun, S. Fu, L. Wang, J. Hu and Y. Gao, "Internet Financial Fraud Detection Based on a Distributed Big Data Approach With Node2vec," in IEEE Access, vol. 9, pp. 43378-43386, 2021, doi: 10.1109/ACCESS.2021.3062467.

[23]. A. Jung and P. H. J. Nardelli, "An Information-Theoretic Approach to Personalized Explainable Machine Learning," in IEEE Signal Processing Letters, vol. 27, pp. 825-829, 2020, doi: 10.1109/LSP.2020.2993176.

[24]. Y. Li et al., "Automated Anomaly Detection via Curiosity-Guided Search and Self-Imitation Learning," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 6, pp. 2365-2377, June 2022, doi: 10.1109/TNNLS.2021.3105636.

[25]. W. A. Al-Khater, S. Al-Maadeed, A. A. Ahmed, A. S. Sadiq and M. K. Khan, "Comprehensive Review of Cybercrime Detection Techniques," in IEEE Access, vol. 8, pp. 137293-137311, 2020, doi: 10.1109/ACCESS.2020.3011259.

[26]. S. Yu, X. Li, X. Zhang and H. Wang, "The OCS-SVM: An Objective-Cost-Sensitive SVM With Sample-Based Misclassification Cost Invariance," in IEEE Access, vol. 7, pp. 118931-118942, 2019, doi: 10.1109/ACCESS.2019.2933437.

[27]. Z. Yuan, H. Chen, T. Li, X. Zhang and B. Sang, "Multigranulation Relative Entropy-Based Mixed Attribute Outlier Detection in Neighborhood Systems," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 52, no. 8, pp. 5175-5187, Aug. 2022, doi: 10.1109/TSMC.2021.3119119.

[28]. P. Lertwanitrot and A. Ngaopitakkul, "Discriminating Between Capacitor Bank Faults and External Faults for an Unbalanced Current Protection Relay Using DWT," in IEEE Access, vol. 8, pp. 180022-180044, 2020, doi: 10.1109/ACCESS.2020.3026744.

[29]. T. Patcharoen and A. Ngaopitakkul, "Transient Inrush and Fault Current Signal Extraction Using Discrete Wavelet Transform for Detection and Classification in Shunt Capacitor Banks," in IEEE Transactions on Industry Applications, vol. 56, no. 2, pp. 1226-1239, March-April 2020, doi: 10.1109/TIA.2019.2963251.

[30]. M. Zhong, T. Xue, Y. Song, S. X. Ding and E. L. Ding, "Parity Space Vector Machine Approach to Robust Fault Detection for Linear Discrete-Time Systems," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 7, pp. 4251-4261, July 2021, doi: 10.1109/TSMC.2019.2930805.