Data Analysis for Fraud Detection in Finance

Raghad Almutairi
ralmuta@gmu.edu

Abhishek Godavarthi
agodavar@gmu.edu

Arthi Reddy Kotha
akotha2@gmu.edu

***Abstract*—Credit card use is not always the best way to use for payments, but the most demonstrable payment mode is through the credit card for both offline a s w ell a s f or online payments, which can result in deficit o f f unds. A s t he online shopping is booming it helps in rendering the cashless payment modes. It can be used at shopping's, paying rent, paying utilities bill, internet bill, travel and transportation, entertainment, food. Using for all these things there is a chance of fraud transactions for a credit card, hence there is more risk. There are many types of fraudulent detections most of the banks and institutions are preferring fraud detection applications.it has become very hard to find o ut t he f raud d etections, A fter t he t ransaction is done there is a chance of detecting fraudulent transactions in the manual business processing system. In real time the bunco transactions are done with real transactions, but it seems not to be sufficient for detecting [1]. Machine learning and data science both are playing a very important role in identifying the fraud detections. This study uses data science and machine learning for detecting the fraud detection to demonstrate various modellings. The problem enables the transactions of the previously done transaction data.**

***Index Terms*—Credit card, Banking Services, Fraud detection, Cashless Payments**

The dataset has been collected from a research collaboration of Worldline and the Machine Learning Group of ULB (Université Libre de Bruxelles) on big data mining and fraud detection. For statistical and visual analytics - Python has been used. The results would be visualized using various methodologies through these tools. The original dataset will been cleaned if necessary for accurate analysis of the data.

Fig. 1. Types of Fraud

I. Introduction

Since 1950, when credit cards were first f ound, a nd until this day, fraudsters have been competing to devise new ways to steal personal information to reach people's credit cards and obtain money. Fraud is a deceptive action done intentionally to give the culprit access to the victim's sensitive information illegally to gain an advantage or money, and it is also known as a "false representation of facts" [15]. Fraud often happens when fraudsters know about information the victim needs, which naturally leads the victim to believe the scam thinking of it as a reliable source or person [15]. There are many different types of fraud. Many people fall victim to the most common fraud types: health care fraud, business and investment fraud, mail fraud, insurance fraud, consumer fraud, charity and disaster fraud, social media and internet fraud, and financial f raud [16].

This research paper focuses on financial f raud a nd credit card fraud. Credit card fraud is unauthorized access to someone else's credit card information to buy something online or illegally transfer an amount of money to another bank account [17]. Moreover, credit card fraud also has different types and methods, as shown in the following graph.

With the popularity of e-commerce, most companies and institutions offer their products and services online through websites. The online-shopping websites require consumers to create accounts and add personal information, including credit card information such as card number, date, and name. Furthermore, unsecured websites may expose consumers to credit card fraud or identity fraud because these websites are considered easy targets for hackers and fraudsters. Even though most websites now use a two-step verification method to increase security and verify the user's identity, fraudsters still find ways to defraud users while shopping online using various methods.

Consequently, companies and organizations with unsecured or less secured websites eventually lose their customers as customers will not trust sharing sensitive information on an un-

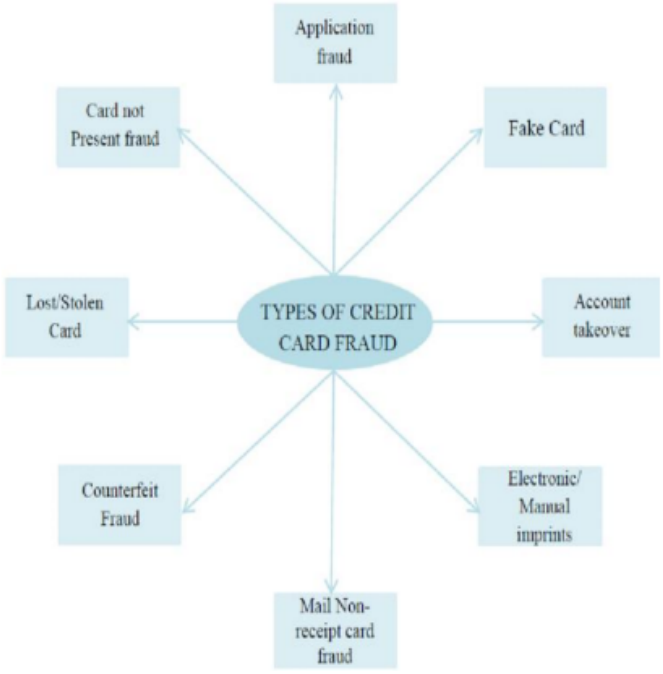

Fig. 2. Types of Credit Card Fraud

secured site. For example, Banks must maintain confidentiality and keep customers' personal information and transactions secured. Many banks have lost their customers and closed their doors after failing to protect customers' financial privacy and information. In addition, as banks and other institutions suffer consecutive losses due to fraud, other factors will be affected, such as the economic reputation.

This research paper shows the standard terms and highlights the key statistics of credit card fraud detection. The proposals made in this paper in terms of savings and time efficiency are the beneficial attributes. The techniques used in this paper reduce credit card fraud, but there is also a chance that it misguides the people to be misclassified as fraudulent. Sometimes there was some ethics in the banking and the complexity and efficiency of obtaining the money. It deals with cases that involve criminal cases, which could be challenging to identify sometimes.

Credit card fraud detection identifies any attempt to purchase or transfer from a credit card made by a fraudulent and stops the transaction instead of processing it [19].

There are two techniques most often used to detect credit card fraud. The first one is fraud analysis which is analyzing the fraudulent transactions of credit cards, and the second one is anomaly detection which is used to detect the anomalies based upon the previous data of transactions [5].

To detect credit card fraud using machine learning is to use and analyze data to investigate the habits and methods of fraudsters and build a model that helps detect and reduce fraudulent transactions [5]. To create a model to detect fraud, the data science team will need to collect the credit card users' data, such as the habit patterns, area, product types, amount, and spending, and then use the data and information to discover the fraudsters' behavioral patterns [5].

The different techniques used by various companies for detecting fraud transactions are Artificial Neural Networks, Hidden Markov Model, Naïve Bayes, the KNN classifier, and Genetic Algorithm [5]. Online user analysis based on natural language processing models can improve credit fraud detection significantly [2]–[9] [10], [11] [**?**], [12]–[23] [24]–[37].

II. Problem Statement/objectives

The increase in credit card fraud leads card users to lose money and companies to lose consumers. Credit card fraud directly affects consumer spending, which can also damage the economy. Every time consumer spending drops off, the prices drop, and the economic growth slows down [1].

In the United States, credit card fraud has significantly increased in the past few years as the number of reports went up from around 130K in 2017, to almost 400K report in 2020 [19].

Credit card fraud can happen in two different ways. The first way is to steal information or identity and open a new bank account with someone else's identity to get their money while avoiding all responsibility.

The second way is accessing an existing account. To succussed in credit card fraud, the fraudsters will have to access personal information such as full name, email address, ID number, credit card number, and other sensitive information, which can also lead to Identity theft.

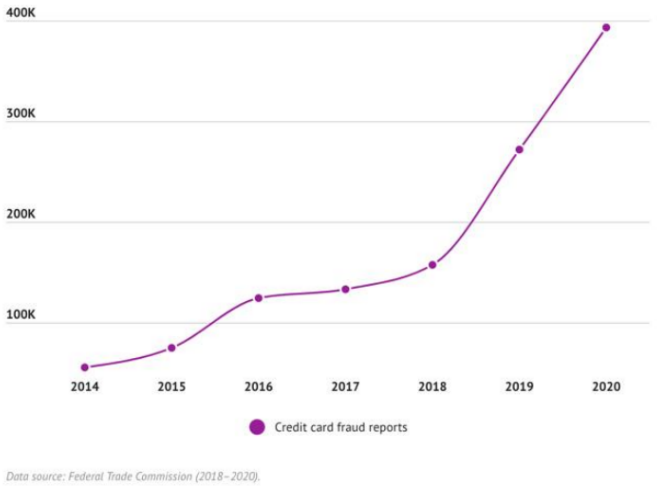

Fig. 3. Credit Card Fraud Reports

Identity theft is another related issue as fraudsters steal people's personal information and use it to create a credit card with the stolen identity [2].

The goal of Identity thieves here is to steal money, and this can be a big issue as the number of people in debt will sharply increase, causing more people to lose jobs, lose homes, and go to prison, which will also affect the economy as a result. Covid-19's impact is one of the main reasons identity thefts has become more common in the last two years. According

to Daly (2021), about 1.4 million reports of identity theft were received in the United States in 2020, and a report from Aite Group stated that identity theft caused losses estimated at $712.4 billion in 2020 [2]. In this research, we want to prevent this problem and reduce the losses as much as possible [9].

Another problem is credit card scams. These days scammers are creative and use many ways to scam people into stealing their credit card information. Some scammers use phone calls to pretend to be bank employees, and others may send links to their victims, leading to a fake login page with the same original page designed to steal people's login information [3]. Scammers use many ways that most people are not aware of, especially elderly and young people. Moreover, many hackers also use users' lack of awareness to steal and use their personal information.

Recognizing identity and credit card fraud transactions and identifying fraudsters' behavior is essential for every institution, including banks, to preserve the users' information and secure the institution's database and system. With the increase in the number of customers and technology development, creating a model that will help detect fraudulent transactions automatically and as accurately as possible is substantial.

Some of the important questions that need to be answered are : how to deal with imbalanced data and misclassified data and how to propose a conceptual framework from an almost anonymized real-world credit card dataset.

III. Rationale and Significance of The Study

Given the rapid digital transformation that businesses are witnessing, the need to improve credit card services and products' security is even more essential and urgent. Such a need requires increasingly secure information systems in organizations [5]. Businesses need to ensure that their information technology infrastructure is upgraded with the latest that emerging and advanced technologies offer. In order to achieve this, a holistic understanding of the business is needed, and ongoing efforts to recognize any gaps and areas for improvement to safeguard the company's financial health. Because a substantial amount of financial dealings such as credit card processing are performed online, the transactional credit card data can help constantly learn to detect anomalous activities.

IV. Literature Survey

Numerous research projects have been published about credit card fraud detection since the 90s and until now. Fraud transactions are one of the illegal activities going on in the world today. Fraud is also considered a crime nationally and internationally. As discussed above, this research will check all the possible ways to determine the fraud transaction targets and what steps to take forward. Since the 19th century, people have faced problems solving fraud probabilities and handling them. However, in the 20th century, a new algorithm was introduced, and that is the Bayesian algorithm that was used in different ways to solve the issues, also called frequentist statistics. There was also some deep learning proposal and topologies derived from the detections. The two types of random forests that are used for fraud transaction detection are normal and abnormal transactions.

In the investigation, there were several techniques used for fraudulent transactions and are also used for comparing the credit data. When presenting the crime data of the credit card there are certain issues faced by the data mining as well as non-data mining thefts [10].

The invoice bills and the payments are also issued to the investigator to investigate the credit card detections, known as no cash mobile applications. An alarm shows when your transaction is confirmed as a fraudulent or malicious transaction [10].

The machine learning-based approach to financial fraud detection also uses the mobile application. In this research we use the methods surveyed by Kambar et al. in study [**?**] on the machine learning techniques for mobile malware detection. Mobile payment fraud is identified in the growing issues of credit cards. The decision tree is also used for fraudulent detection in the credit card system. It proposed the sampling process and the selection process for the feature used with the large amount of the transaction data, which has high accuracy in mobile payment. Four modules are being used: data collection, data preprocessing, feature extraction, and evaluation model. Random forest selects the best feature for the data resulting from the better model for the users [13]. The advancements of the techniques with machine learning give us complete information and new techniques to learn.

The machine learning-based approach to financial fraud detection also uses the mobile application. Mobile payment fraud is identified in the growing issues of credit cards. The decision tree is also used for fraudulent detection in the credit card system. It proposed the sampling process and the selection process for the feature used with the large amount of the transaction data, which has high accuracy in mobile payment. Four modules are being used: data collection, data preprocessing, feature extraction, and evaluation model. Random forest selects the best feature for the data resulting from the better model for the users [13]. The advancements of the techniques with machine learning give us complete information and new techniques to learn. Credit card fraud detection using machine learning is done by developing new classifications and regressions [10].

This is an exceptionally applicable issue that requests the consideration of networks, for example, AI and information science, where the answer for this issue can be mechanized. This issue is complicated according to the point of view of learning, as different factors like class unevenness portray it. The number of legitimate exchanges far dwarfs false ones. Additionally, the exchange designs frequently change their real properties throughout the natural process of everything working out. Extortion is the unlawful or criminal trickiness expected to bring about monetary or individual advantage. A purposeful demonstration is an illegal rule or strategy used to achieve an unapproved monetary advantage. Various written works relating to irregularity or extortion recognition in this

space have been distributed as of now and are accessible for public use [14].

Many models have been recommended for precise fraud detection. For instance, we can consider the brain network proposed by Ghosh and Reilly, which is prepared on a large sample of named Visa exchanges. These transactions contain an assortment of misrepresentation cases, such as lost cards, stolen cards, application extortion, email misrepresentation, etc. Training on an assortment of information makes the model resistant to almost any sort of misrepresentation. Subsequently, the quality or assortment of data matters more than the amount of bulk. There were numerous different methodologies before, and there will be numerous later, and there is still a ton of scope for this field as the cheats keep on being unavoidable [14].

Some research papers covered different methods for detecting financial fraud, and others included different learning techniques. These research papers were used to develop and improve fraud detection systems and algorithms. The previous work discussed two methods for machine learning: deep learning and traditional machine learning.

Deep learning is machine learning, but the main difference is the performance. Deep learning "uses a programmable neural network that enables machines to make accurate decisions without help from humans" (Grieve, 2020, para. 5) [4].

fraud detection [5]. In the study, deep learning has been used to solve complex problems. The research paper includes studying deep learning methods for credit card fraud detection issues and comparing this method's performance with other machine learning algorithms on three different financial datasets. This research shows that deep learning methods had better performance than traditional machine learning models and that the results can be effectively used in the real world.

Another study by Zanin, Romance, Moral, and Criado in 2017 discussed how to detect credit card fraud through parenclitic network analysis. This study presented the first complex network classification algorithm to detect criminal cases in an actual card transaction dataset. This research shows how including features from the network data representation can improve the obtained result by a standard neural network-based classification algorithm [6].

A comprehensive survey conducted by Clifton Phua and his associates have revealed that techniques employed in this domain include data mining applications, automated fraud detection, adversarial detection. In another paper, Suman, Research Scholar, GJUS&T at Hisar HCE presented techniques like Supervised and Unsupervised Learning for credit card fraud detection. Even though these methods and algorithms fetched an unexpected success in some areas, they failed to provide a permanent and consistent solution to fraud detection [25].

A similar research domain was presented by Wen Fang YU and Na Wang where they used Outlier mining, Outlier detection mining and Distance sum algorithms to accurately predict fraudulent transaction in an emulation experiment of credit card transaction data set of one certain commercial bank. Outlier mining is a field of data mining which is basically used in monetary and internet fields. It deals with detecting objects that are detached from the main system i.e. the transactions that aren't genuine. They have taken attributes of customer's behaviour and based on the value of those attributes they've calculated that distance between the observed value of that attribute and its predetermined value [25].

Unconventional techniques such as hybrid data mining/complex network classification algorithm is able to perceive illegal instances in an actual card transaction data set, based on network reconstruction algorithm that allows creating representations of the deviation of one instance from a reference group have proved efficient typically on medium sized online transaction.

Multiple Supervised and Semi-Supervised machine learning techniques are used for fraud detection [20], but we aim is to overcome three main challenges with card frauds related dataset i.e., strong class imbalance, the inclusion of labelled and unlabelled samples, and to increase the ability to process a large number of transactions.

Differet Supervised machine learning algorithms [21] lie Decision Trees, Naive Bayes Classification, Least Squares Regression, Logistic Regression and SVM are used to detect fraudulent transactions in real time datasets.Two methods under random forests [22] are used to train the behavioural features of normal and abnormal transactions. They are Random-tree-based random forest and CART-based. Even though random forest obtains good results on small set data, there are still some problems in case of imbalanced data. The future work will focus on solving the above-mentioned problem. The algorithm of the random forest itself should be improved.

Performance of Logistic Regression, K-Nearest Neighbour, and Naïve Bayes are analysed on highly skewed credit card fraud data where Research is carried out on examining meta-classifiers and meta-learning approaches in handling highly imbalanced credit card fraud data.

Through supervised learning methods can be used there may fail at certain cases of detecting the fraud cases.A model of deep Auto-encoder and restricted Boltzmann machine (RBM) [23] that can construct normal transactions to find anomalies from normal patterns. Not only that a hybrid method is developed with a combination of Adaboost and Majority Voting methods [24].

A. Modes of Fraud Detection usually used

The algorithms that are used for the model detections of credit card are as follows, they are Random Forest Classifier, Decision tree, Support Vector machine.

Random Forest Classifier: It is an inconsequential classifier among all the classifiers. It is an essential module of the classification. For the better evaluation and performance depends on several factors that includes depth, maximum bins, impurity [5].

Decision Tree: There is a test node which is represented with the root not and each node. Occurrences, Target Attribute,

Attributes List are their elements used for planning tree which is called as leaf node [5].

Support Vector Machine: Based on the other algorithms this is best used for fraud detection. In this all the information or data is combined into one category and SVM is one of the form used to separate the data which was converted or separated by the different classes [5].

V. Dataset Review

The dataset has been collected from a research collaboration of Université Libre de Bruxelles on big data mining and fraud detection. This study uses a sample of data to detect fraudulent credit card transactions and prevent any credit card charges that take place without the card owner's permission. Moreover, utilizing the data to build a model that is simple but fast enough to stop the fraud in time as well as report it to the relevant authorities.

The dataset contains transactions made using credit cards by European cardholders in September 2013 and it shows that out of 284,807 transactions that were made in two days, there were 492 frauds and only 0.17% of the transactions were fraudulent. Unfortunately, the dataset does not provide more background about the data, but we will try to search about more data in other datasets the future. The credit card detection model we will be building will include Anomaly Detection which is the process of finding rare events, items, or actions that is considered suspicious to detect fraudsters. The anomaly detection will be applied by using machine learning where both normal and anomalous samples will be included to make the model predict future actions. We want the model to be trained by using reports and feedbacks to find the probability of fraud to give an alert [8].

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, they cannot provide the original features and more background information about the data. Features V1, V2, ... V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'.

Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example dependant cost sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise[8].

VI. Proposed Approach

The technique we are using in our research to detect fraud is the fraud analyzing technique, where we analyze fraud transactions using a model to detect fraudulent behaviors. The data obtained from the dataset will be processed and analyzed properly, cleaned, and visualized. We will use the data later to try and build several machine learning models in order to find the best solution.

As shown in Figure 5, these are the steps to follow to accomplish the model.

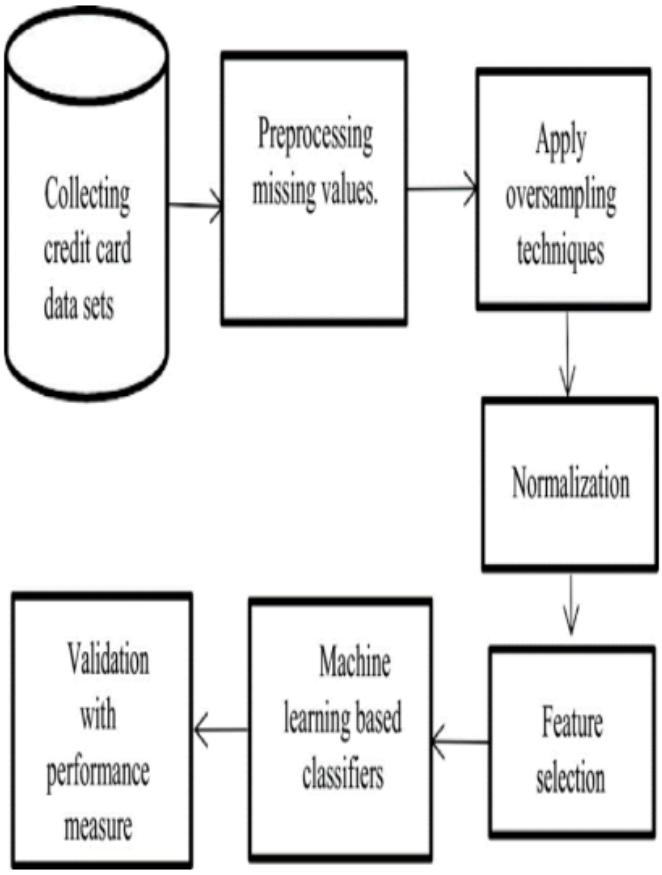

Fig. 4. Approach

After collecting, cleaning, and visualizing our data, the next step is to start building the model.

The process will start by building a prototype , using multiple open-source machine learning libraries to facilitate the comparison process. First, after installing the libraries, all the necessary files will be imported, the dataset will be loaded, and the classifications will be set up[12].

Next, the comparison process begins as we start creating the best possible model, saving, loading, finalizing, and deploying the model.

A. Data Preprocessing

Data preprocessing involves transforming the raw data into the well-formed data sets so that the analysis can be performed in a better way.

Data preprocessing is done in steps that follows as formatting, cleaning, sampling. Formatting is used for putting the data in a way that it will be suitable for the work. The data and the formatting changes according to the needs of the people.

The most recommended or most used format is the CSV format. Data cleaning is the main and important procedure in the field of data science. It does a major part of the work. That is 80% of the work is done. Sampling is the technique for analyzing all the subsets with the whole large datasets, which

gives us a better result for understanding and the pattern of the data in the integrated way.

Data is represented differently and are kept together with the help of data integration. When the data is large, or the dataset is huge then the databases can become very slow to work. To replace the data with the raw values within given intervals reduces the number of values by diving that in the given rage.

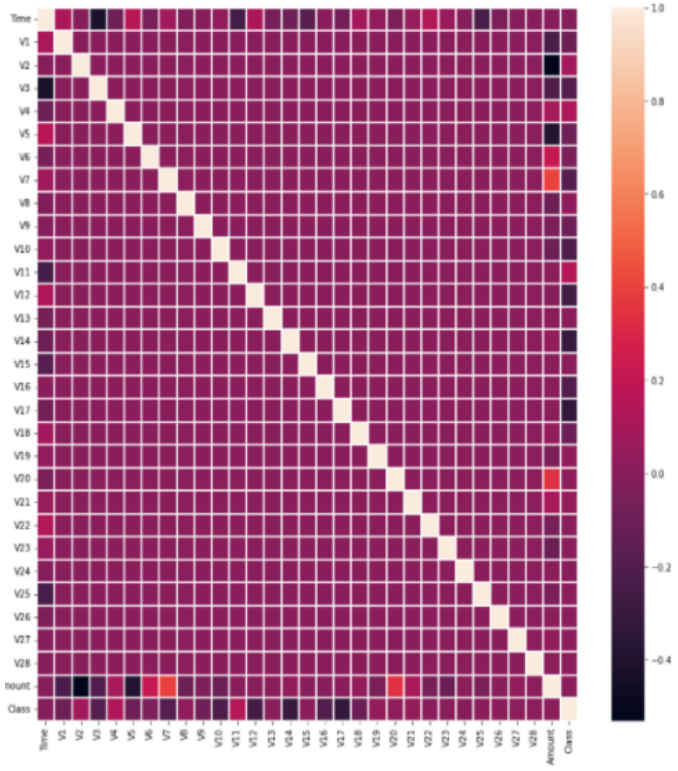

Fig. 5. Correlation Matrix

There is no notable correlation between features V1-V28.There are certain correlations between some of these features and Time (inverse correlation with V3). Features and Amount (direct correlation with V7 and V20, inverse correlation with V1 and V5).

Results on the data preprocessing showed that there were 798 missing values distributed in the 'V22' and 'V23' columns and 1076 duplicated values. These two rows were removed as the dropped data did not significantly impact the result.

The outliers were inlcuded as thet might affect the accuracy of the model as the model may mislabel for the extreme cases in further analysis.

After the data preprocessing 2,82,783 observations remained with 282,356 data for the negative class and 427 data for the positive class.

The result indicates that the dataset is highly imbalanced.

VII. Results

Based on the above two distributions 'V3,' 'V8,' 'V13,' 'V15,' 'V19,' 'V21' and 'V24' have a similar distribution curve.

Based on the above correlation we can say that Features only having a weak correlation to Class. As part of dimension reduction, principal component analysis has been used. We chose 27 components for 95% variance.

A. Classfication Models

Three classfication models have been used to evaluate the estimators for the model. Hyperparameter tuning was used to maximize the accuracy of the model.

1) K-Nearest Neighbor model: Using this model the best score was 0.915 during the training process. The testing process gave the model accuracy of 96.7%.

The number of true positives and true negatives were 47 and 98 respectively. The number of false positives and false negatives were 2 and 3 respectively.

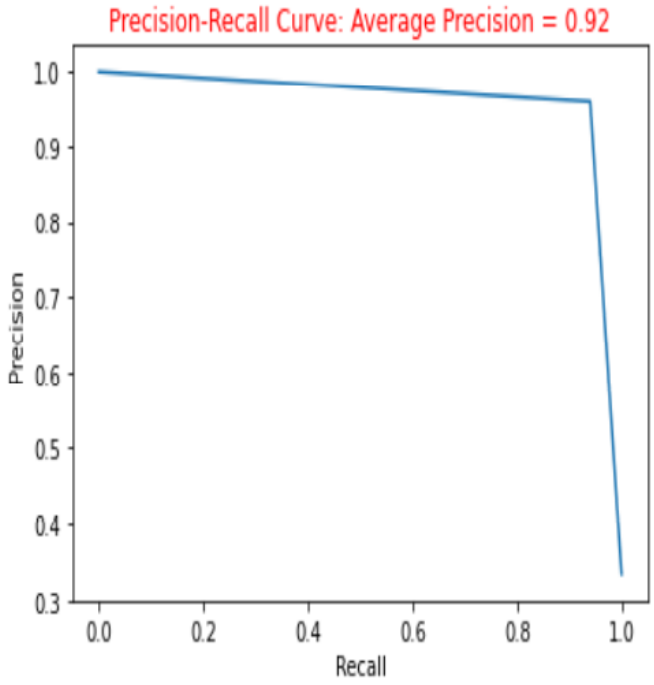

Fig. 6. Precision recall curve KNN model

2) Random Forest Model: Hyperparameter tuning with 5-fold cross-validation was used to maximize the accuracy of the model and to to choose the best estimators.

Using this model the best score was 0.94 during the training process. The testing process gave the model accuracy of 94%.

The number of true positives and true negatives were 48 and 90 respectively. The number of false positives and false negatives were 10 and 2 respectively.

3) XGBoost Model: Using this model the best score was 0.939 during the training process. The testing process gave the model accuracy of 92%.

Hyperparameter tuning with 5-fold cross-validation was used to maximize the accuracy of the model and to to choose the best estimators.

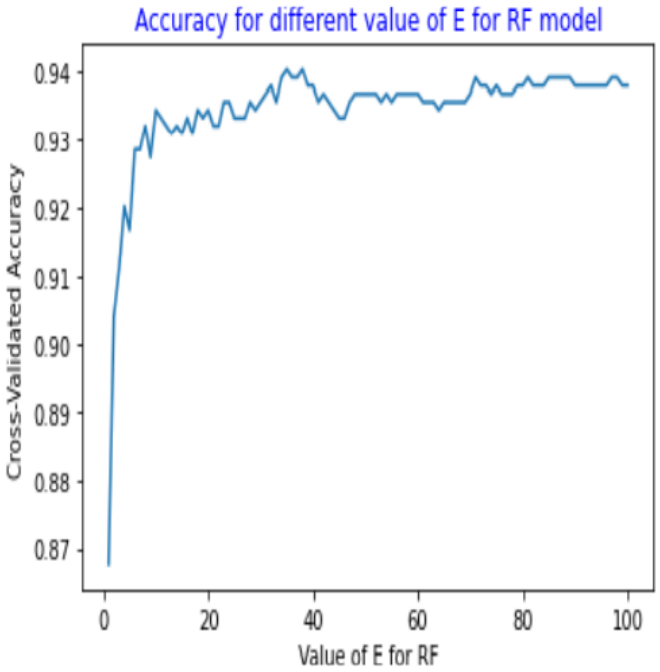

Fig. 7. Accuracy using Random Forest

The number of true positives and true negatives were 49 and 89 respectively. The number of false positives and false negatives were 11 and 1 respectively.

VIII. Conclusion

Credit card fraud detection methods have gained the popularity in the past decade with the evolution of the statistical models. Fraud detection is an essential task for the merchant bank which involve the customer and their bank and therefore there is a need for improvement using the combination of one more algorithms.

In this research, we have cleaned and preprocessed the data, checked for data unbalancing and could identify the relationship between various features.

Three Machine learning algorithm models have been implemented to train and test the data.

We started with KNN, for which we obtained an accuracy of 96.7% . We followed with Random Forest, which resulted in a comparatively lower accuracy of 94%.We then followed with an XB Boost, which resulted in the lowest accuracy of 92%.

Weighting the false positives and false negatives is also important if we care more about detecting a particular group of fraudulent transactions. Based on the results XGBoost classification can be used as it had the least False negatives (type 2) [17], [21]

References

[1] K. Amadeo. What you buy every day drives u.s. economic growth. the balance. (accessed February 23, 2022). [Online]. Available: https://www.thebalance.com/consumer-spending-definition-and-determinants-3305917

[2] M. Heidari and J. H. Jones, "Using bert to extract topic-independent sentiment features for social media bot detection," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, 2020, pp. 0542–0547.

[3] M. Heidari, J. H. Jones, and O. Uzuner, "Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter," in *2020 International Conference on Data Mining Workshops (ICDMW)*, 2020, pp. 480–487.

[4] M. Heidari and S. Rafatirad, "Semantic convolutional neural network model for safe business investment by using bert," in *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2020, pp. 1–6.

[5] M. Heidari, J. H. J. Jones, and O. Uzuner, "An empirical study of machine learning algorithms for social media bot detection," in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 2021, pp. 1–5.

[6] M. Heidari and S. Rafatirad, "Bidirectional transformer based on online text-based information to implement convolutional neural network model for secure business investment," in *2020 IEEE International Symposium on Technology and Society (ISTAS)*, 2020, pp. 322–329.

[7] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, "Emotion detection of textual data: An interdisciplinary survey," in *2021 IEEE World AI IoT Congress (AIIoT)*, 2021, pp. 0255–0261.

[8] D. E. Knuth, *Seminumerical Algorithms*. Addison-Wesley, 1973c1981.

[9] ——, *Seminumerical Algorithms*, 2nd ed., ser. The Art of Computer Programming. Reading, Massachusetts: Addison-Wesley, 10 Jan. 1973c1981, vol. 2, this is a full BOOK entry.

[10] M. Heidari and S. Rafatirad, "Using transfer learning approach to implement convolutional neural network model to recommend airline tickets by using online reviews," in *2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA*, 2020, pp. 1–6.

[11] S. Zad, M. Heidari, J. H. Jones, and O. Uzuner, "A survey on concept-level sentiment analysis techniques of textual data," in *2021 IEEE World AI IoT Congress (AIIoT)*, 2021, pp. 0285–0291.

[12] L. Daly. Identity theft and credit card fraud statistics for 2021. (accessed February 23, 2022). [Online]. Available: https://www.fool.com/the-ascent/research/identity-theft-credit-card-fraud-statistics/

[13] B. Myers. Credit card fraud and scams: How to avoid both. (accessed February 23, 2022). [Online]. Available: https://www.fool.com/the-ascent/credit-cards/scams-fraud-how-avoid/

[14] P. Grieve. Deep learning vs. machine learning: What's the difference? (accessed February 23, 2022). [Online]. Available: https://www.fool.com/the-ascent/credit-cards/scams-fraud-how-avoid/

[15] T. T. Nguyen, H. Tahir, M. Abdelrazek, and A. Babar, "Deep learning methods for credit card fraud detection," *CoRR*, vol. abs/2012.03754, 2020. [Online]. Available: https://arxiv.org/abs/2012.03754

[16] M. Zanin, M. Romance, S. Moral, and R. Criado, "Credit card fraud detection through parenclitic network analysis," *CoRR*, vol. abs/1706.01953, 2017. [Online]. Available: http://arxiv.org/abs/1706.01953

[17] C. Oden. Design and implementation of a credit card fraud detection system. (accessed February 23, 2022)(n.d). [Online]. Available: https://www.projecttopics.org/design-and-implementation-of-a-credit-card-fraud-detection-system.html

[18] GeeksforGeeks. Ml — credit card fraud detection. (accessed February 23, 2022). [Online]. Available: https://www.geeksforgeeks.org/ml-credit-card-fraud-detection/

[19] Saloni and M. Rout, "Analysis and comparison of credit card fraud detection using machine learning," in *Advances in Electronics, Communication and Computing*, P. K. Mallick, A. K. Bhoi, G.-S. Chae, and K. Kalita, Eds. Singapore: Springer Singapore, 2021, pp. 33–40.

[20] M. Thirunavukkarasu, A. Nimisha, and A. Jyothsna, "Credit card fraud detection through parenclitic network analysis," *International Journal of Computer Science and Mobile Computing*, vol. 10, no. 4, pp. 71–79, 2021.

[21] techopedia. Data preprocessing. (accessed February 23, 2022). [Online]. Available: https://www.techopedia.com/definition/14650/data-preprocessing

[22] @amankrsharma3. Automating the machine learning pipeline for credit card fraud detection. (accessed February 23, 2022). [Online]. Available: https://www.geeksforgeeks.org/automating-the-machine-learning-pipeline-for-credit-card-fraud-detection/

[23] akanksha singh and A. Singh. Ieee-cis fraud detection. (accessed April 18, 2022). [Online]. Available: https://www.academia.edu/44187814/IEEE_CIS_

[24] M. Heidari, S. Zad, B. Berlin, and S. Rafatirad, "Ontology creation model based on attention mechanism for a specific business domain," in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 2021, pp. 1–5.

[25] M. Heidari, S. Zad, and S. Rafatirad, "Ensemble of supervised and unsupervised learning models to predict a profitable business decision," in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 2021, pp. 1–6.

[26] P. Hajibabaee, M. Malekzadeh, M. Ahmadi, M. Heidari, A. Esmaeilzadeh, R. Abdolazimi, and J. H. J. Jones, "Offensive language detection on social media based on text classification," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, 2022, pp. 0092–0098.

[27] S. Zad, M. Heidari, P. Hajibabaee, and M. Malekzadeh, "A survey of deep learning methods on semantic similarity and sentence modeling," in *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2021, pp. 0466–0472.

[28] M. Heidari, J. H. J. Jones, and O. Uzuner, "Online user profiling to detect social bots on twitter," 2022. [Online]. Available: https://arxiv.org/abs/2203.05966

[29] M. Heidari, S. Zad, P. Hajibabaee, M. Malekzadeh, S. HekmatiAthar, O. Uzuner, and J. H. Jones, "Bert model for fake news detection based on social bot activities in the covid-19 pandemic," in *2021 IEEE 12th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, 2021, pp. 0103–0109.

[30] P. Hajibabaee, M. Malekzadeh, M. Heidari, S. Zad, O. Uzuner, and J. H. Jones, "An empirical study of the graphsage and word2vec algorithms for graph multiclass classification," in *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2021, pp. 0515–0522.

[31] M. Malekzadeh, P. Hajibabaee, M. Heidari, S. Zad, O. Uzuner, and J. H. Jones, "Review of graph neural network in text classification," in *2021 IEEE 12th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, 2021, pp. 0084–0091.

[32] R. Abdolazimi, M. Heidari, A. Esmaeilzadeh, and H. Naderi, "Mapreduce preprocess of big graphs for rapid connected components detection," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, 2022, pp. 0112–0118.

[33] A. Esmaeilzadeh, M. Heidari, R. Abdolazimi, P. Hajibabaee, and M. Malekzadeh, "Efficient large scale nlp feature engineering with apache spark," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, 2022, pp. 0274–0280.

[34] S. Rafatirad and M. Heidari, "An exhaustive analysis of lazy vs. eager learning methods for real-estate property investment," 2019. [Online]. Available: https://openreview.net/forum?id=r1ge8sCqFX

[35] M. Malekzadeh, P. Hajibabaee, M. Heidari, and B. Berlin, "Review of deep learning methods for automated sleep staging," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, 2022, pp. 0080–0086.

[36] M. Heidari and J. H. J. Jones, "Bert model for social media bot detection," 2022. [Online]. Available: http://hdl.handle.net/1920/12756

[37] M. Heidari, "Nlp approach for social media bot detection(fake identity detection) to increase security and trust in online platforms," 2022.