Educational Administration: Theory and Practice
2024,30(4), 1188-1199
ISSN:2148-2403
https://kuey.net/

Research Article

Real Time Streaming Of Credit Card Synthetic Transactions And Imbalanced Pattern Analyses Of Extracted Dataset

Rinku[1*] , Ashutosh Kumar Dubey[2], Sushil Kumar Narang[3], Neha Kishore[4]

[1] * Chitkara University School of Computer Applications, Chitkara University, Himachal Pradesh, India, rinku.cse@chitkarauniversity.edu.in
[2]Chitkara University School of Engineering and Technology, Chitkara University, Himachal Pradesh, India, ashutosh.dubey@chitkara.edu.in
[3]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India, sushilk.narang@chitkara.edu.in
[4]University Institute of Engineering and Technology Maharaja Agrasen University, Himachal Pradesh, India, nehakishore.garg@gmail.com

Citation: Rinku et Al. (2024), Real Time Streaming Of Credit Card Synthetic Transactions And Imbalanced Pattern Analyses Of Extracted Dataset ..*Educational Administration: Theory And Practice*, *30*(4), 1188-1199
Doi: 10.53555/kuey.v30i4.1635

ARTICLE INFO

ABSTRACT

Simulation of the real-time transactions can be achieved through the backend servers while Credit Card expenses online or offline mode. Credit Card transactions simulator generates the data in real time and generate the synthetic imbalanced dataset that can be used for different purpose along with the fraud detection. The synthetic data sets doesn't contain the personal or private data of any person or any disclosure of any legal transactions. So, this dataset can be used for design, development of fraud detection models. Synthetic dataset can be extracted from the realtime transactions placed through the application where the Credit Card transactions are placed. Further this dataset is synthetic so it's fast and secure and easy to acquire and customizable for any experimentation. This paper intends to design and develop a synthetic dataset generator Credit Card Transaction Simulator for Credit Card transactions in real time and generates data that usefully approximates the relevant aspects of the real data transactions with the imbalanced patterns.

Keywords: Synthetic Dataset, Simulator, Imbalanced Dataset, Credit Card Dataset, Credit Card Transactions, Credit Card Fraud, Credit Card Fraud Detection

Introduction

Credit card fraud detection is an essential task in the financial industry to prevent fraudulent transactions and protect customers from financial losses(Iqbal and Amin, 2023). Machine learning techniques have been widely used in credit card fraud detection due to their ability to learn from historical data and make predictions on new transactions(Almazroi and Ayub, 2023). However, credit card fraud detection poses a significant challenge due to the imbalanced nature of the data, where fraudulent transactions are rare events compared to legitimate transactions. In an imbalanced dataset, a machine learning model tends to predict the majority class more often, leading to poor performance on the minority class(Huang *et al.*, 2023). To address this issue, several techniques have been proposed, such as oversampling, undersampling, and cost-sensitive learning. However, the choice of hyperparameters in the classification model also plays a crucial role in determining the model's performance on imbalanced data

Data is available universally in various forms like text, images, audio, video etc. Everything that can be recorded would be consider as data(Likhitha and Mohan, 2017). Data words comes from the plural word datum that sights the piece of information. The data can be categories into various forms, classifications or types and stored into the various databases. Some of them are structured, semi-structured , timeseries and operational data etc. (Alkhatib *et al.*, 2021)**.** Real time data is most important along with the data stored into the various databases. Generally, data generated from various sources and stored into the database to extract the useful information. An extracted part of data from the database considered as the dataset (Barddal *et al.*,

2020).Digitally a collection of data records for information processing is called dataset. In the today's scenario, as the e-commerce becomes more popular among the peoples, online transactions increased rapidly(Taha and Malebary, 2020). The payment methods using Credit Cards grown up instead of using cash in their daily life for normal type of payments. Payments using a Credit Card facilitates the users to track the payments and their status of payments(Tanouz *et al.*, 2021a). Many companies and institutions shifted their business from cash to cashless with the help of modern technology. Credit Card provides the payment method to pay any type of billing, rent, booking etc. at very fast transaction mode(Choi and Lee, 2018). All these types of transactions completed at the bank server in the form of batch processing. A database or log is generated along with the successful and failure transaction with the definition of the reason of failure(Olowookere and Adewale, 2020).

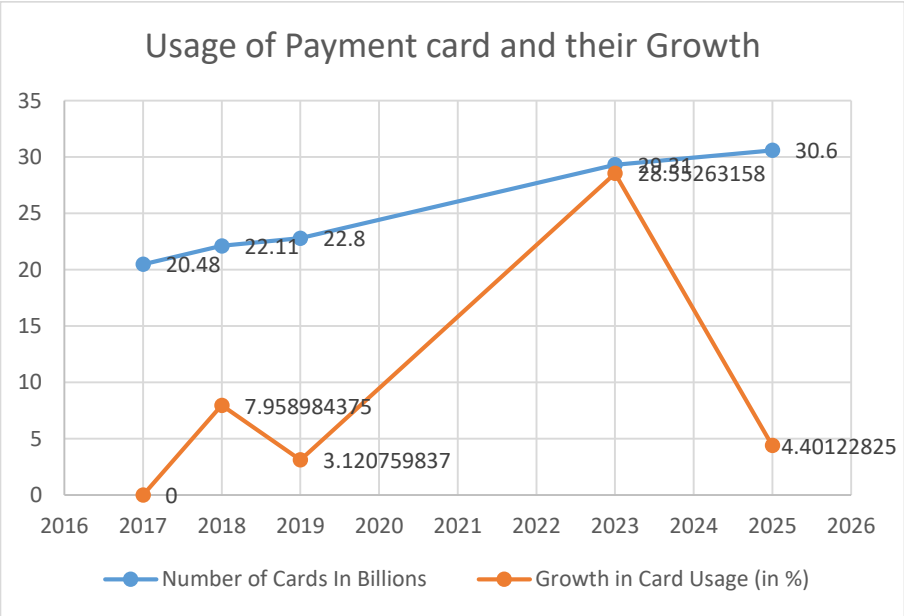

Figure-1: growth of card usage (in %) and number of cards (in billions)

The increase in the number of payment card will be a great with relative to the growth in the offline card usage. (Kennedy et al., 2023)Most of the time the payment will be done through the online mode through the UPI of the bank account or the credit card. As per the Nilson report Payment cards usable only for purchases from select retailers, fuel stations, medical and dental facilities and other private label locations generated $937.75 billion in total volume in 2020, down from $975.22 billion the prior year(El Naby, El-Din Hemdan and El-Sayed, 2021). Losses to fraud on these cards was $0.67 billion in 2020, up from $0.66 billion the prior year. Private label cards were tied to 2.34% of global fraud losses in 2020 ATM cash advances and withdrawals initiated by global brand cards are counted in this report in the total volume attributed to those credit and debit cards (Almazroi and Ayub, 2023) . Another $1.717 trillion in cash volume at ATMs occurred outside of the global brand card networks in 2020. This was down from $1.908 trillion in 2019. Fraud losses from ATM transactions of $1.40 billion in 2020 was down from $1.55 billion in 2019. ATM fraud accounted for 4.90% of global fraud losses in 2020. In 2030, when total volume on all payment cards is expected to reach $79.140 trillion, fraud losses are projected to be $49.32 billion, equal to 6.23 cents per $100. In the U.S., total volume in 2030 is projected at $18.953 trillion and fraud losses are expected to be $17.00 billion, equal to 8.97 cents per $100. Over the next 10 years, card industry losses to fraud will collectively amount to $408.50 billion. Gross fraud losses incurred by issuers of credit, debit and prepaid cards were $18.69 billion in 2020 compare to $19.59 billion in 2019. Card issuers accounted for 65.40% of gross losses to fraud worldwide in 2020. The other 34.60% of fraud losses, which equaled $9.89 billion, were incurred by merchants, ATM acquirers and merchant acquirers. In 2019, that group experienced $9.06 billion in losses to fraud.
The discussion in paper divided into different sections; the first section discusses about the process of a credit card transactions on server and merchant side. The next section of the paper discussed about the various dataset available for the analysis. The third section of the paper shows the analysis of the dataset pattern and the imbalancing present in the dataset. The section four discussed about the limitation of the existing datasets. The next section of the paper discussed about the experimental setup of the proposed credit card transaction simulator (CCTS). The result and discussion part display the sample dataset extracted from real time streaming, transactions happed on the server and the testing of the API using software. The next sections discussed about the conclusion followed by future work.

I. Fundamental Process of Credit Card transactions During Purchase

Credit Card transaction takes place online as well as in offline modes. Many things happen when a Credit Card is use for the purchase from any merchant(Karthikeyan, Govindarajan and Vijayakumar, 2023). Most of the

process proceeds behind the scenes and the user will not able to see the actual processing. A few entities are involved into every Credit Card transaction. In a scenario a customer wants to purchase an item from any merchant in online or offline mode. After selection of the item consumer present his/her Credit Card for the payment. (Wang *et al.*, 2018).
The merchant sells the services or the goods generally. The merchant's bank sends Credit Card transactions for the payment approval. The Credit Card payment network is a cooperation between the merchant bank and the Credit Card issuer(Tanouz *et al.*, 2021b).The Credit Card bank issuer gives the approval and complete the transaction as display in the figure-1.

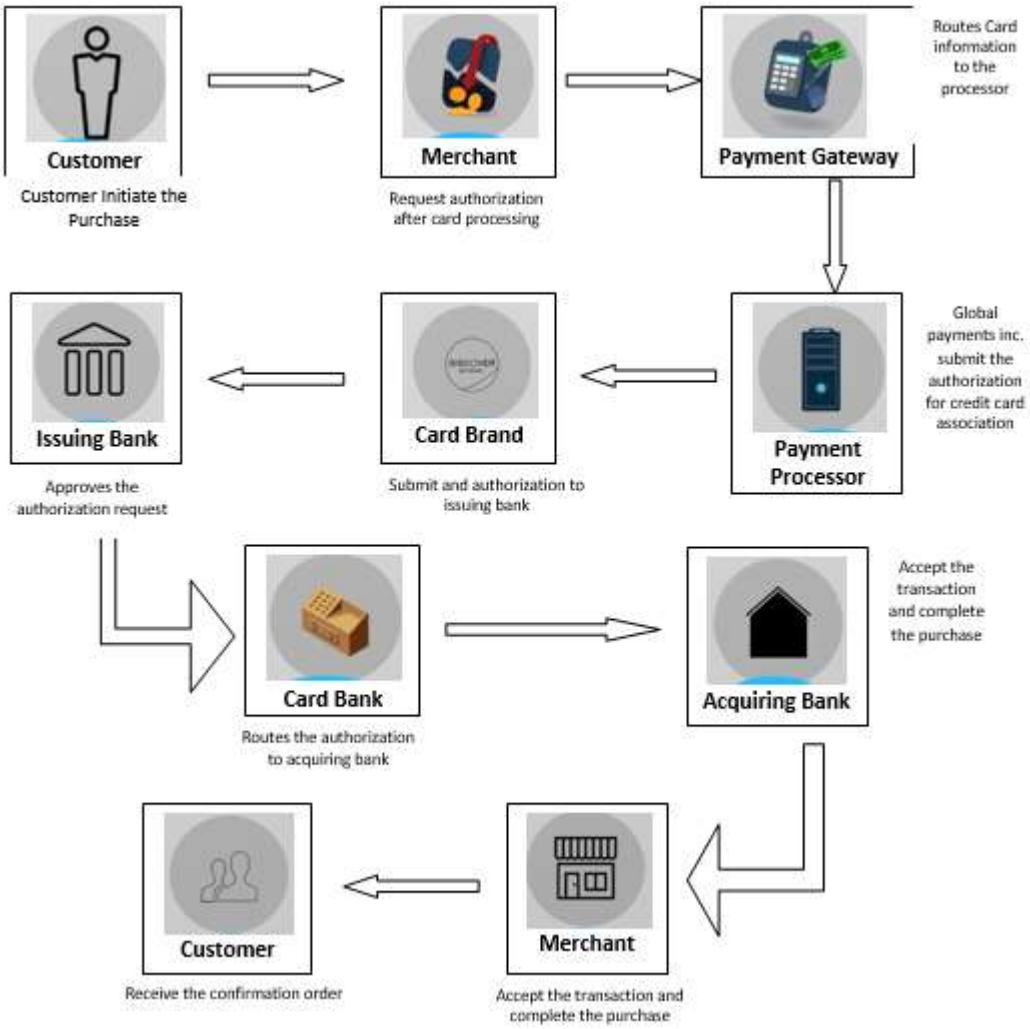

Figure-1: Fundamental process flow for Credit Card transaction during purchase

II. Credit Card Transactions Datasets

A set of data would be considered as the Credit Card dataset generated from the transactions of the Credit Card in online or offline mode(Warghade, Desai and Patil, 2020). There are different payment methods are available using that the user can do the payments and all these transactions will be recorded on the servers. Time to time the data can be used for the research purpose. Credit Card transaction dataset contains various fields' viz. transaction date, transaction time, amount, transaction-id, location, type of Card, Card owner name etc. Fraudulent transactions are the major problem for e-commerce business today(Fang, Zhang and Huang, 2019).Fraud detection have the biggest problem of classification to find the truly and a fraudulent transaction. The data sets generated from the Credit Card transactions generally they are highly imbalanced (El Naby, El-Din Hemdan and El-Sayed, 2021). Privacy is the most important concern that's why in the sensitive transactions fields names are changed.

Importance of Credit Card datasets

A fraudulent transaction places on the web server of the bank. Finding the frauds into the transactions exhibits the great importance for any financial institution offering or using Credit Cards (Sadgali, Sael and Benabbou, 2018). This becomes essential for organizations to detect the fraud and stop them before happening for security reasons also. Credit Card datasets provides the dummy or old transactions on that experiments can be done to

find the frauds into the transactional data (Univerzitet u Istočnom Sarajevu. Faculty of Electrical Engineering *et al.*, 2019). The rationale of Datasets is to avoid straight communicating with the database using simple SQL statements.

In most of the software application direct communication with the database can be created by the developer to access the data from the database. The datasets reside completely in memory, making them useful for temporary tables. Because datasets are RAM based, they are extremely fast. The datasets store their data in a very efficient manner, making them resource friendly (Najadat *et al.*, 2020).The datasets can automatically calculate averages, subtotals, and totals over a group of records.

Existing Datasets of Credit Card Transaction

Datasets are available on different websites. The websites having names mentioned owns the copyrights on the data and authorizes its reproductions(Nghiem, Thu and Nghiem, 2018; Alharbi *et al.*, 2022).

- Kaggle [A dataset Repository]
- UCI Machine Learning Repository [Verified Dataset Repository]
- Econometric Analysis Book by William H. Greene
- Credit scoring and its applications Book by Lyn C. Thomas
- Credit Risk Analytics Book by Harald, Daniel and Bart
- Lending Club
- PAKDD 2009 Data Mining Competition, organized by NeuroTech Ltd. and Center for Informatics of the Federal University of Pernambuco

Datasets of the Credit Card transactions are available online on the different websites. Some others websites using the same dataset available on kaggle or UCI machine learning repository(Manlangit *et al.*, 2018). Some example sites are mentioned as below:

- https://www.kaggle.com/datasets
- https://archive.ics.uci.edu/ml/index.php
- https://datahub.io/machine-learning/CreditCard
- https://data.world/vlad/Credit-Card-fraud-detection
- http://eforexcel.com/wp/downloads-17-sample-csv-files-data-sets-for-testing-Credit-Card/
- https://catalog.data.gov/dataset?tags=Credit-Card

III. Analysis of Existing Dataset

Dataset analysis is very important to design the simulator. A random and most popular dataset of Credit Card is used for analysis purpose. This dataset is available on the kaggle website and used by the various researchers(Ren *et al.*, 2019; Mahdi, Pardede and Ali, 2021). The dataset consists of transactions made by the Credit Card in the September 2013 by the person's European cardholder. This dataset has transactions that were occurred in the period of two days, it contains 492 frauds out of 284,807 transactions. The dataset is highly imbalanced , due to the presence of the fraudulent records amount 0.172% of all transactions(Parmar, C. Patel and Savsani, 2020).

This dataset contains the numerical values those are the result of the transformation of the Principal Component Analysis of the data. Regrettably, because of confidentiality issues, dataset did not provide the original features and more background information about the data. Features mentioned as V1, V2, ... V28 are the principal components obtained with PCA, the feature Time and Amount has not been transformed with PCA. Feature in the dataset 'Time' contains the seconds onwards between every transaction and the starting transaction in the dataset. The feature 'Amount' into the dataset shows the transaction Amount, this feature can be used for cost-sensitive learning. Dataset hold a feature 'Class' having value 0 or 1 for non-fraud and fraud transaction (Wen and Yusuf, 2019).

There are some well-known websites who are providing the similar kind of datasets that also can be used for the various purposes of detections of problems in the Credit Card transactions. These datasets are mentioned into the Table-1 are available as open source and can be downloaded any time. Various researchers used these datasets for the implementation of multiple models of detection and resolutions. The table consists of the records of various types of Credit Card transactions, starting and ending date of the transactions along with the number of transactions done during the period.

Table-1: Published dataset on the web(Bayram, Koroglu and Gonen, 2020)

Sr. No	Credit Card Transactions Record set	Starting Transaction Date	Ending Transaction Date	Number of transactions
1	Fraud Train	6/21/2020	12/31/2020	1048575
2	Fraud Test	6/21/2020	12/31/2020	555719

3	Online Retail	12/1/2010	12/9/2011	541910
4	Online Retail	12/1/2010	12/9/2011	541910
5	Online Retail-II	12/1/2009	12/9/2010	525461
6	Credit Card Transaction	10/4/2013	5/26/2015	26052
7	corporate Credit Card transaction	1/5/2015	31/08/2015	3865
8	HDFC Transactions (Data.gov.uk)	1/1/2019	1/4/2019	176

Datasets mentioned into the Table-1 provides the huge number of transactions of Credit Card usage in different domains. These dataset provides the patterns of transaction generation. To understand the pattern of the records generation of a dataset a sample dataset of Online Retail-II is used and some sample is displayed in below graph (Figure-2 and Figure-3) that shows the number of transactions in every second of time interval. The dataset Online Retail-II consists of 525461 records having the details of the transactions in multiple domains.

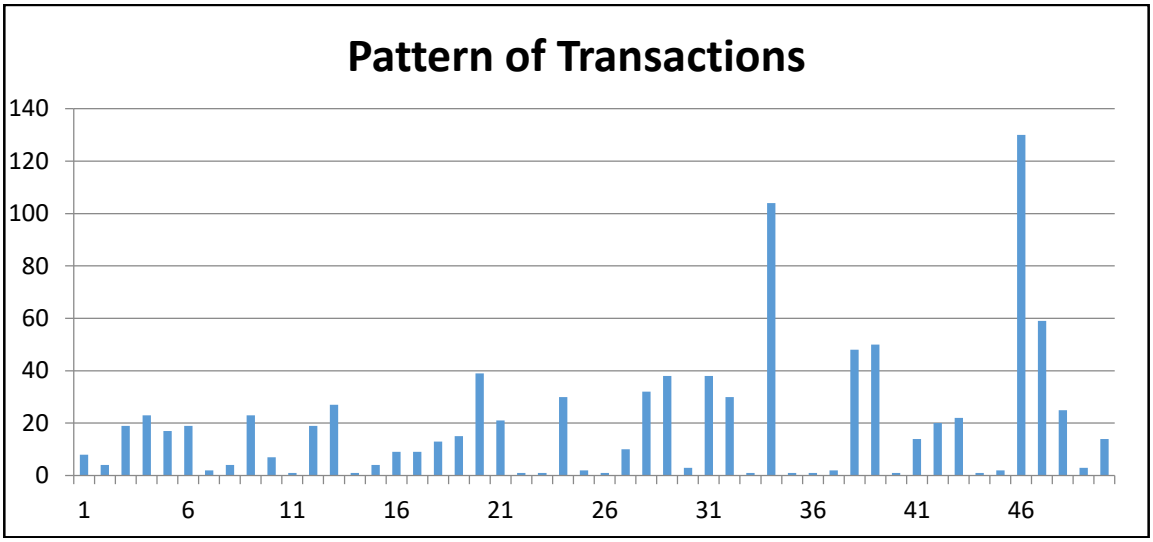

Figure-2: Transactions in unit (in seconds) interval of time Sample-1[Online Retail-II Dataset]

Datasets also give the information about the transactions happened into every second of time. Figure -1 and Figure-2 are displaying the records of consecutive 50 records. The graphs show the data imbalance and the concept drift into the data generation.

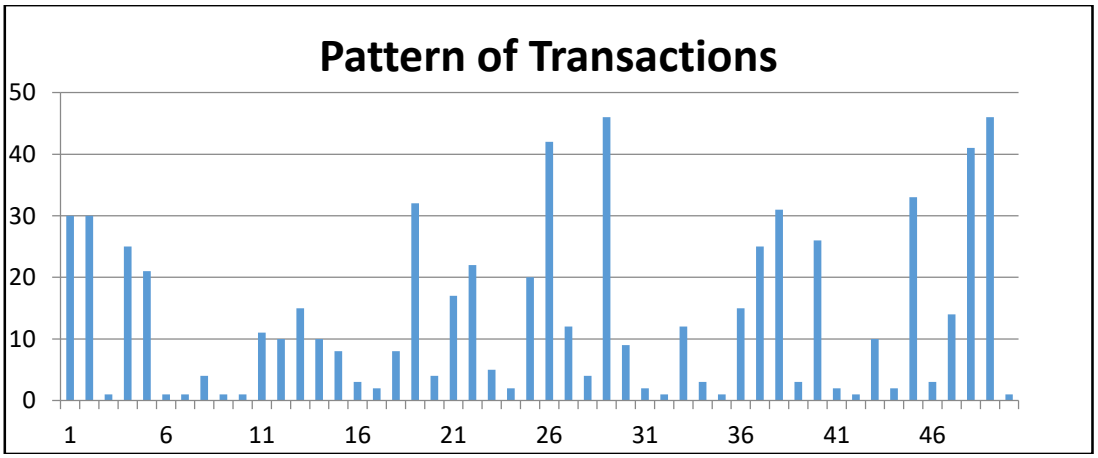

Figure-3: Transactions in unit (in seconds) interval of time Sample-2[Online Retail-II Dataset]

To analysis and understand the pattern of the dataset of Online Retail-II from the total of 525461 records found and 25297 instances are found into the dataset. The findings from the dataset of the online retail regarding the record generation is given into the Table-2.

Parameter	Sample-1	Sample-2
Number of Instances in the sample	50 Instances	50 Instances
Maximum transactions in the sample per second	130 transactions	46 transactions

Table-2: Pattern Analysis of records in dataset [Online Retail-II]

IV. Limitation of the existing Credit Card datasets

Credit Card datasets are available online and can be downloaded anytime and use for multiple number of times(Ren *et al.*, 2019; Wang *et al.*, 2019). Major issues suggested by the various researchers are

- Understanding complex domain
- Cleaning dirty data, usually from multiple sources
- Deciding which analytics method to use, often not straightforward
- Visualizing results for the end-user or decision-maker, who doesn't know much about analytics Some of the other major issues are also having the real-time datasets
-

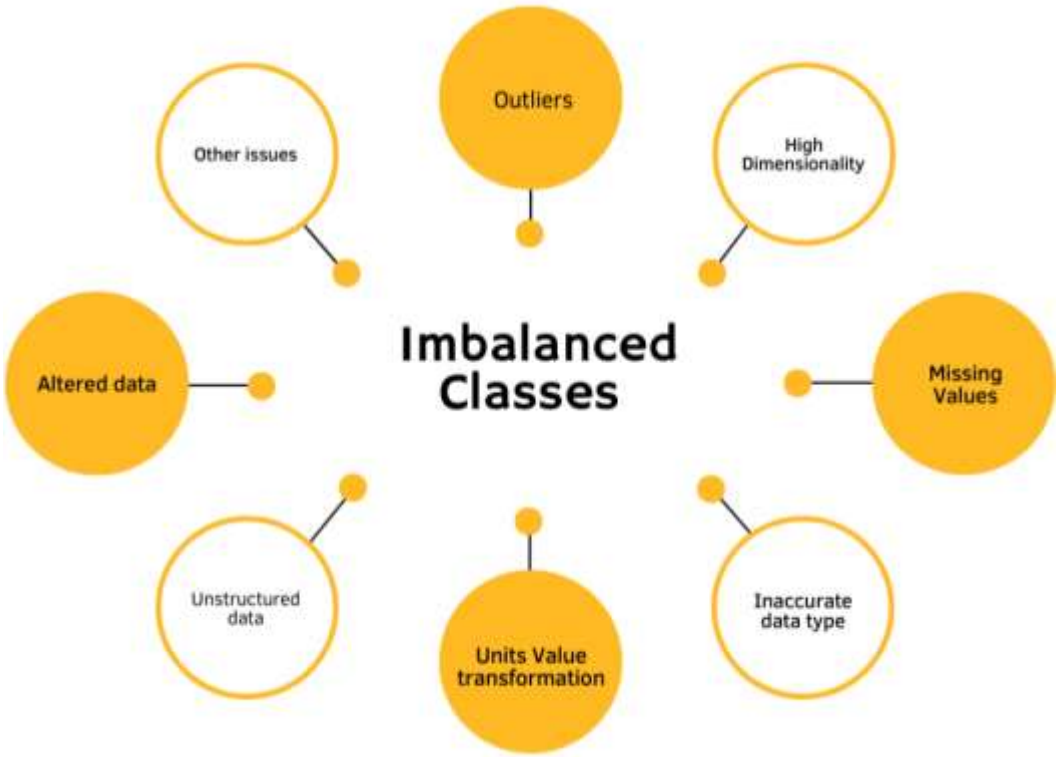

Figure-4: Issues with real-time datasets

The dataset generated in real time generally have imbalanced classes(Vikrant Agaskar *et al.*, 2017; Tanouz *et al.*, 2021a). Certain issues like unstructured-ness, inaccurate data type selection, missing values etc. are found into the datasets.

V. Credit Card Transaction Simulator [CCTS]

The dataset those are published over the web is available for download at any time and can be used for the detection of fraud. But as we know the number of transactions doesn't depend on the time and in a particular time periods the count of transactions might vary. As discussed by researchers that fraudulent transactions share the very less percentage into the pool of transactions of Credit Card. So, the detection of fraud into the real time streaming data is much more beneficial for the organizations and the users.

Modelling of the dataset of Credit Card transactions with respect to the fraudulent records into the ansaction in data streaming requires the real time analysis and learning of the data. The dataset available on the web aren't real time and the learning from them may not as much effective.

A web application name Credit Card Transactions Simulator (CCTS) implemented into the spring-boot and eclipse. CCTS simulates the Credit Card transactions over the web server. CCTS mainly implemented for following reasons:

• This CCTS provides the real time streaming data of Credit Card transactions

- The model or algorithms have input data that is generating in real time and no limit of the records that can be used for the learning of the model.
- The parameters in the CCTS are customizable and there are enough records can be downloaded as per need.
- Dataset can be extracted at any point of time and it is purely synthetic.

Experimental Setup

To implement the simulator, firstly the previous datasets are analyzed. Based on the analysis of the various samples, their transaction in unit time (in seconds) and the patterns of the transactions, CCTS is developed in Spring-boot 4.0 and Java Development Kit version 8.0. Application API is tested on the POSTMAN software application. Application is developed using Eclipse IDE. HTML and CSS is used to design and develop the page. JSP help to make the page dynamic. The application is deployed with the help of TOMCAT server 9.0 that is bundled along with the spring-boot workspace.

To implement a CCTS a web application based on micro-services is developed using Springboot Framework. This helps in creating fully production ready environment that is completely configurable with code into its codebase. The application API was tested using POSTMAN that takes the data into the JSON format. This application deployed using Apache Tomcat Server 9.0 to host on the web. This application gets called on various systems and devices and generates the transactions. All the generated transaction goes to the server and stored into the database MySQL. Simultaneously the admin on the server can take the transactional data at any point of time as they are stored into the database.

The application runs on various locations and generates the data and send to the server. On the server they create a message queue. As this queue is provided to the Apache kafka server as input it provides the data in real time. As data is generated from any source using the application the Apache kafka server just produce the record synchronously as shown in figure 5. given below.

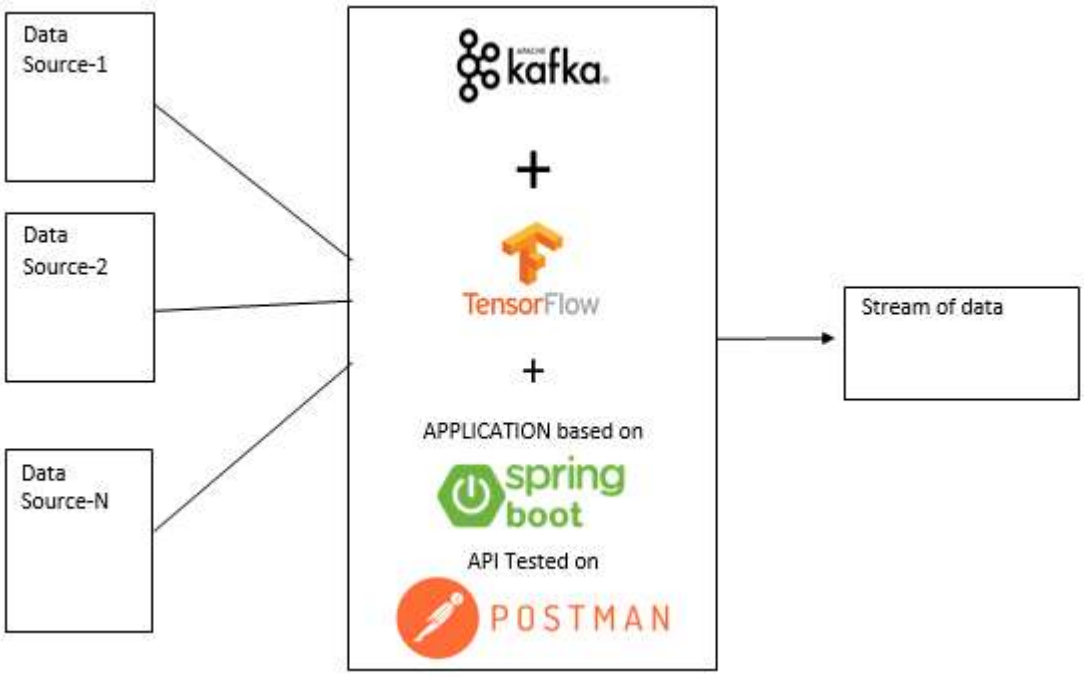

Figure-5: Real time transaction streaming

The application contains various pre-defined and user-defined algorithms to get the customized result. In the customization the features and the patterns those are learned from the previously available datasets were considered. The synthetic dataset is extracted from a real time environment where an application is running that accepts the transactional data from various sources. Apache Kafka and TensorFlow accepts the input from the multiple sources and devices and generate a transactional data stream. This application CCTS (Credit Card Transactional System) is customizable and desired data amount can be extracted with customizable features of data. A synthetic dataset can be extracted as per the need of the application. The figure given below depict the process of data stream generation

Throughout this work, a dataset is explored containing all the transactions issued by Credit Card. The Credit Card transactions contain various features(Fatima *et al.*, 2021). These features describe the Card-holder, the merchant and the transaction that connects them in the bipartite graph of transactions. The Card-holder attributes are describing the Card-holder or the Card: pan-id, age, gender, card expiry, province code, district code, ins code, zip, city, country, term-miduid, term-mcc, term-category, tx-amount, tx-datetime, tx-time, tx-

ecom, tx-3dsecure, tx-emv, tx-currency, card number, transaction id, card type, card-authentication, card cvv, card entry mode, tx fraud etc. are the various attributes of the Credit Card transactions used in various Credit Card datasets available on the web(VOICAN, 2021).
JSP application is created in the Spring-boot using eclipse IDE (Integrated Development Environment) that is able to generate the transaction continuously. Various API (Application Programming Interface) are created in Java to get the transaction_id, Card_no, transaction date etc.

VI. Result and Discussion

The CCTS application is designed and developed using spring boot and API (application programming interface) tested using POSTMAN. The application is embedded with Kafka and TensorFlow to generate the credit card transactional data stream. The basic need of the CCTS to generate the synthetic dataset that behaves like an original transactional dataset. Here CCTS provide the environment where transactions a placed at different places and the web application responds accordingly. It behaves same as the original transaction takes place. A number of transactions is generated from the different client devices and all the transactions will be commute to the server. Many of the customers purchase the services or goods from the various platforms using a Credit Card.
The below figure shows the CCTS transactions happening on the server.

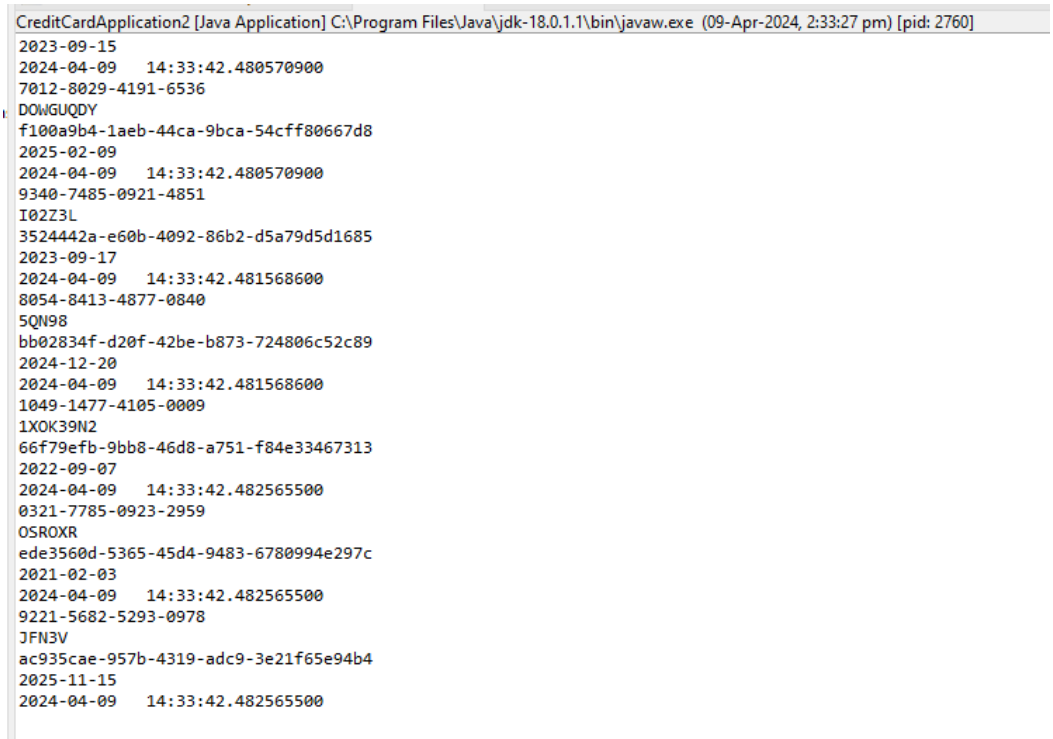

Figure-7: Snapshot of transactions on CCTS server

These API were tested on the POSTMAN software application as mentioned in figure-8

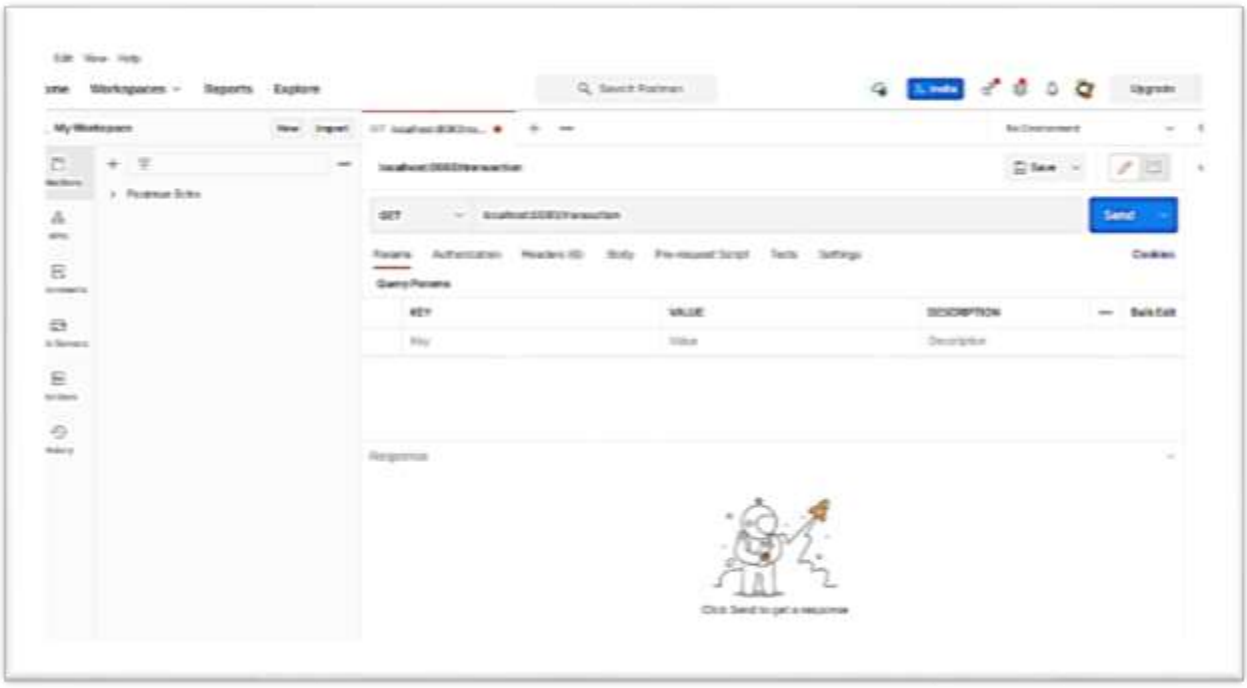

Figure-8: CCTS API Testing on POSTMAN

The merchant's objective is to serve the customers and commit the payment that result into the generation of a data set. In virtual environment the transaction is always between merchant and customer with the payment gateway in the middle. Buying from another customer or selling articles to another merchant is not embedded in CCTS.The simulator generates the transactions as real-time output as output in the figure-9 in the form

Sr No	Pan-ID	Age	Gender	Card Number	Card expiry	Transactions ID	----------
1							

Figure-9: snapshot of the sample dataset output from CCTS

CCTS provide the customized prospect to get the data and the number of features in real time. Dataset can be downloaded in a customized was as per the requirement of the user. The extracted sample given in figure-10

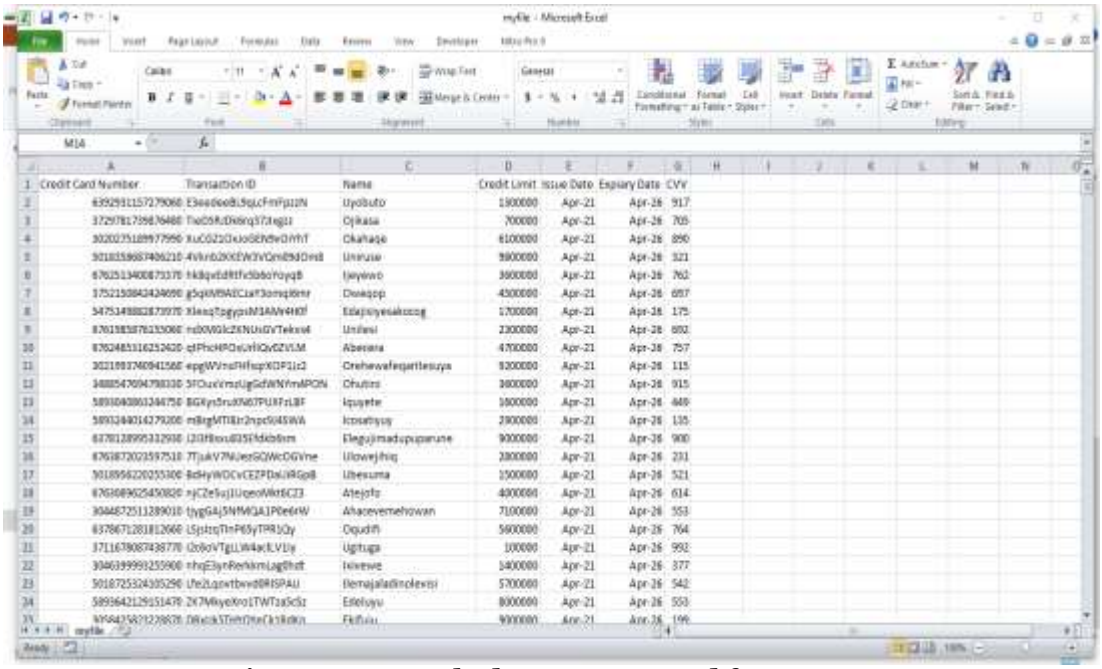

Figure-10: Sample dataset extracted from CCTS

In figure-11 the transactions patterns have been displayed, it contains the number of transactions per seconds placed in the CCTS server.

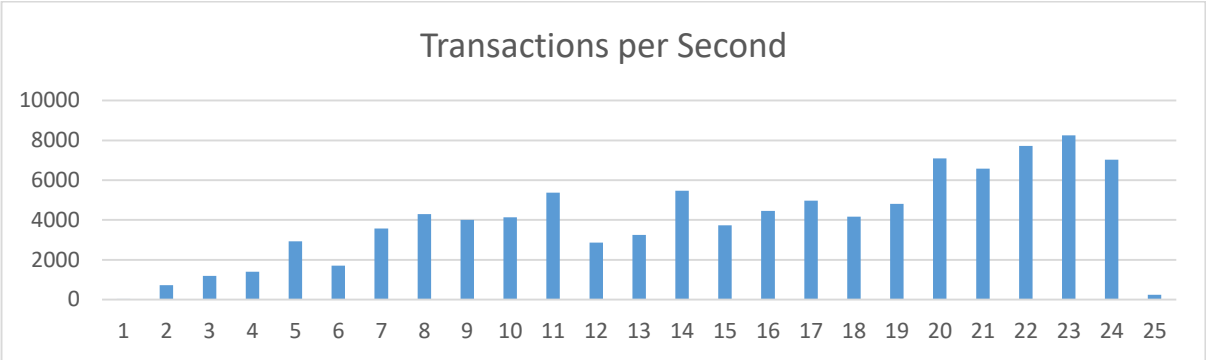

Figure-11: Transactions in unit (in seconds) interval of time

Here five samples of transactions have taken from the CCTS and analyzed on the different parameters like time elapsed, average rate of transactions, median and deviations in the transactions rate. These sample details are given below in table-3.

Table-3: Sample transactional analysis of credit card data

Samples	*Time Elapsed	Minimum transactions (in unit second)	Maximum transaction (in unit second)	Average rate of transaction	Median	Standard deviation
Sample-1 (1 Lakh)	19.7	804	8366	4000	4137	2253.39
Sample-2 (2 Lakhs)	24.8	1024	9377	7692.31	8739	2412.99
Sample-3 (3 Lakhs)	42	667	8380	7184.86	7647	1494.97
Sample-4 (4 lakhs)	52	381	9437	7547.14	7792	1739.27
Sample-5 (5 lakhs)	66	436	9585	8065.52	8507	1834.1

The given table-3 and the figure-12 showed that the dataset samples extracted from the real time streaming using CCTS is highly imbalanced and can be utilized for the implementations of the various fraud detections models.

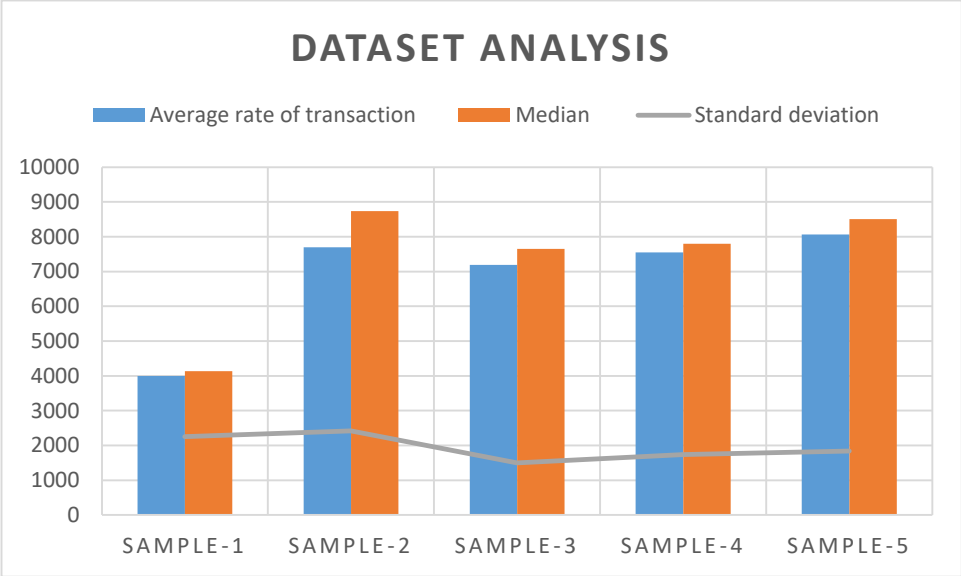

Figure-12: Dataset samples analysis on the parameters

The sample synthetic dataset extracted from the server in the form of a csv file. This dataset is available on given link
https://www.kaggle.com/datasets/rinkuasstprofessor/synthetic-dataset-of-credit-card-transactions.
This synthetic dataset holds ten lakhs' transactions record of the simulated Credit Card transactions.

VII. Conclusion

This is aimed to generate a synthetic dataset of Credit Card transactions with the help of a CCTS. This CCTS is designed and developed into Springboot, Apache-kafka, POSTMAN and deployed on the Tomcat Server. The transactions generated from the web applications goes to the server. In the middle tier Apace Kafka is used to get the real-time transactions. CCTS ran continuously to obtain a distribution that get close enough to be reliable for testing. Since this is a randomised simulation, the values are of course not identical to original data.

CCTS simulates the transactions in the real time data streaming of Credit Card transactions. The CCTS generated data is highly imbalance and contains fraudulent records along with the non-fraudulent records. Generally, the machine learning algorithms operates with the assumptions that the data is static. But in the data streams the transactions generated imbalanced data.

VIII. Future work

The Simulator generated the dataset in real-time and can be downloaded. There are many updation can be done regarding their feature and access mechanism. To use this data in real-time environment it can be implemented using the cloud application like Apache Kafka, Amazon Kinesis firehose; so that it can directly produce its results to a live fraud detection mechanism using a suitable machine learning algorithm.

References

1. Alharbi, A. *et al.* (2022) 'A Novel text2IMG Mechanism of Credit Card Fraud Detection: A Deep Learning Approach', *Electronics (Switzerland)*, 11(5), pp. 1–18. Available at: https://doi.org/10.3390/electronics11050756.
2. Alkhatib, K.I. *et al.* (2021) 'Credit Card Fraud Detection Based on Deep Neural Network Approach', *2021 12th International Conference on Information and Communication Systems, ICICS 2021*, pp. 153–156. Available at: https://doi.org/10.1109/ICICS52457.2021.9464555.
3. Almazroi, A.A. and Ayub, N. (2023) 'Online Payment Fraud Detection Model Using Machine Learning Techniques', *IEEE Access*, 11(November), pp. 137188–137203. Available at: https://doi.org/10.1109/ACCESS.2023.3339226.
4. Barddal, J.P. *et al.* (2020) 'Lessons learned from data stream classification applied to credit scoring', *Expert Systems with Applications*, 162(March), p. 113899. Available at: https://doi.org/10.1016/j.eswa.2020.113899.
5. Bayram, B., Koroglu, B. and Gonen, M. (2020) 'Improving Fraud Detection and Concept Drift Adaptation in Credit Card Transactions Using Incremental Gradient Boosting Trees', *Proceedings - 19th IEEE International Conference on Machine Learning and Applications, ICMLA 2020*, pp. 545–550. Available at: https://doi.org/10.1109/ICMLA51294.2020.00091.
6. Choi, D. and Lee, K. (2018) 'An Artificial Intelligence Approach to Financial Fraud Detection under IoT Environment: A Survey and Implementation', *Security and Communication Networks*, 2018. Available at: https://doi.org/10.1155/2018/5483472.
7. Fang, Y., Zhang, Y. and Huang, C. (2019) 'Credit card fraud detection based on machine learning', *Computers, Materials and Continua*, 61(1), pp. 185–195. Available at: https://doi.org/10.32604/cmc.2019.06144.
8. Fatima, E.B. *et al.* (2021) 'Minimizing the Overlapping Degree to Improve Class-Imbalanced Learning under Sparse Feature Selection: Application to Fraud Detection', *IEEE Access*, 9, pp. 28101–28110. Available at: https://doi.org/10.1109/ACCESS.2021.3056285.
9. Huang, Y. *et al.* (2023) 'A Novel Unsupervised Outlier Detection Algorithm Based on Mutual Information and Reduced Spectral Clustering', *Electronics (Switzerland)*, 12(23), pp. 1–12. Available at: https://doi.org/10.3390/electronics12234864.
10. Iqbal, A. and Amin, R. (2023) 'Time Series Forecasting and Anomaly Detection Using Deep Learning', *Computers and Chemical Engineering*, p. 108560. Available at: https://doi.org/10.1016/j.compchemeng.2023.108560.
11. Karthikeyan, T., Govindarajan, M. and Vijayakumar, V. (2023) 'An effective fraud detection using competitive swarm optimization based deep neural network', *Measurement: Sensors*, 27(March), p. 100793. Available at: https://doi.org/10.1016/j.measen.2023.100793.
12. Kennedy, R.K.L. *et al.* (2023) 'Iterative cleaning and learning of big highly-imbalanced fraud data using unsupervised learning', *Journal of Big Data*, 10(1). Available at: https://doi.org/10.1186/s40537-023-00750-3.
13. Likhitha, M. and Mohan, Y.R. (2017) 'Spark Streaming-Real time stream processing in credit card fraud detection', 6(4), pp. 157–161.
14. Mahdi, O.A., Pardede, E. and Ali, N. (2021) 'Kappa as drift detector in data stream mining', *Procedia Computer Science*, 184(2019), pp. 314–321. Available at: https://doi.org/10.1016/j.procs.2021.03.040.
15. Manlangit, S. *et al.* (2018) *An Efficient Method for Detecting Fraudulent Transactions Using Classification Algorithms on an Anonymized Credit Card Data Set*, *Advances in Intelligent Systems and Computing*. Springer International Publishing. Available at: https://doi.org/10.1007/978-3-319-76348-4_41.
16. El Naby, A.A., El-Din Hemdan, E. and El-Sayed, A. (2021) 'Deep Learning Approach for Credit Card Fraud Detection', pp. 1–5. Available at: https://doi.org/10.1109/iceem52022.2021.9480639.
17. Najadat, H. *et al.* (2020) 'Credit Card Fraud Detection Based on Machine and Deep Learning', *2020 11th International Conference on Information and Communication Systems, ICICS 2020*, (Section IX), pp.

204–208. Available at: https://doi.org/10.1109/ICICS49469.2020.239524.
18. Nghiem, L.T., Thu, T.T. and Nghiem, T.T. (2018) 'MASI: Moving to adaptive samples in imbalanced credit card dataset for classification', *2018 IEEE International Conference on Innovative Research and Development, ICIRD 2018*, (May), pp. 1–5. Available at: https://doi.org/10.1109/ICIRD.2018.8376315.
19. Olowookere, T.A. and Adewale, O.S. (2020) 'A framework for detecting credit card fraud with cost-sensitive meta-learning ensemble approach', *Scientific African*, 8, p. e00464. Available at: https://doi.org/10.1016/j.sciaf.2020.e00464.
20. Parmar, J., C. Patel, A. and Savsani, M. (2020) 'Credit Card Fraud Detection Framework - A Machine Learning Perspective', *International Journal of Scientific Research in Science and Technology*, pp. 431–435. Available at: https://doi.org/10.32628/ijsrst207671.
21. Ren, S. *et al.* (2019) 'Selection-based resampling ensemble algorithm for nonstationary imbalanced stream data learning', *Knowledge-Based Systems*, 163, pp. 705–722. Available at: https://doi.org/10.1016/j.knosys.2018.09.032.
22. Sadgali, I., Sael, N. and Benabbou, F. (2018) 'Detection and prevention of credit card fraud: State of art', *MCCSIS 2018 - Multi Conference on Computer Science and Information Systems; Proceedings of the International Conferences on Big Data Analytics, Data Mining and Computational Intelligence 2018, Theory and Practice in Modern Computing 2018 and Connected Sma*, (March 2019), pp. 129–136.
23. Taha, A.A. and Malebary, S.J. (2020) 'An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine', *IEEE Access*, 8, pp. 25579–25587. Available at: https://doi.org/10.1109/ACCESS.2020.2971354.
24. Tanouz, D. *et al.* (2021a) 'Credit card fraud detection using machine learning', *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, pp. 967–972. Available at: https://doi.org/10.1109/ICICCS51141.2021.9432308.
25. Tanouz, D. *et al.* (2021b) 'Credit card fraud detection using machine learning', *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, (Iciccs), pp. 967–972. Available at: https://doi.org/10.1109/ICICCS51141.2021.9432308.
26. Univerzitet u Istočnom Sarajevu. Faculty of Electrical Engineering *et al.* (2019) '2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH) : proceedings : March 20-21, 2019, Jahorina, East Sarajevo, Republic of Srpska, Bosnia and Herzegovina', *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, (March), pp. 1–5.
27. Vikrant Agaskar, P. *et al.* (2017) 'Unsupervised Learning for Credit Card fraud detection', *International Research Journal of Engineering and Technology*, 4(3), pp. 2395–56.
28. VOICAN, O. (2021) 'Credit Card Fraud Detection using Deep Learning Techniques', *Informatica Economica*, 25(1/2021), pp. 70–85. Available at: https://doi.org/10.24818/issn14531305/25.1.2021.06.
29. Wang, S. *et al.* (2019) 'Learning in the presence of class imbalance and concept drift', *Neurocomputing*, 343, pp. 1–2. Available at: https://doi.org/10.1016/j.neucom.2019.01.080.
30. Wang, Y. *et al.* (2018) 'Privacy Preserving Distributed Deep Learning and Its Application in Credit Card Fraud Detection', *Proceedings - 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications and 12th IEEE International Conference on Big Data Science and Engineering, Trustcom/BigDataSE 2018*, pp. 1070–1078. Available at: https://doi.org/10.1109/TrustCom/BigDataSE.2018.00150.
31. Warghade, S., Desai, S. and Patil, V. (2020) 'Credit Card Fraud Detection from Imbalanced Dataset Using Machine Learning Algorithm', *International Journal of Computer Trends and Technology*, 68(3), pp. 22–28. Available at: https://doi.org/10.14445/22312803/ijctt-v68i3p105.
32. Wen, S.W. and Yusuf, R.M. (2019) 'Predicting Credit Card Fraud on a Imbalanced Data', *International Journal of Data Science and Advanced Analytics*, 1(1), pp. 12–17.