

MTVIS: A FRAMEWORK FOR VISUAL ANALYSIS AND EXPLORATION OF MOBILE MONEY TRANSACTIONS

ZAFFAR AHMED SHAIKH

Faculty of Computing Science and Information Technology, Benazir Bhutto Shaheed University Lyari, Karachi, Pakistan.

ABDUL QAYOOM

School of Information Engineering, Southwest University of Science and Technology, Mianyang, PR China and Department of Computer Science, Lasbela University of Agriculture Water and Marine Sciences, Uthal, Pakistan.

SHAFIQ UR REHMAN

Department of Computer Science, Lasbela University of Agriculture Water and Marine Sciences, Uthal, Pakistan.

ABDULLAH AYUB KHAN*

Faculty of Computing Science and Information Technology, Benazir Bhutto Shaheed University Lyari, Karachi, Pakistan and Department of Computer Science, Sindh Madressatul Islam University, Karachi, Pakistan.

BOCOUM OUSMANE

School of Computer Science and Technology Southwest University of Science and Technology, Mianyang, PR China.

SVETLANA V. MAKAR

Institute of Regional Economy and Interbudgetary Relations, Financial University under the Government of the Russian Federation, Moscow, Russia.

Department of Physical and Socio-Economic Geography, Federal State Budgetary Educational Institution, National Research Mordovia State University, Saransk, Russia.

SERGEY V. SHKODINSKY

Center for Sectoral economics, Financial Research Institute, Moscow, Russia.

Laboratory of industrial policy and economic security, Market Economy Institute of RAS, Moscow, Russia.

TAISIA V. DIANOVA

Department of Economics, Moscow State Institute of International Relations (MGIMO), Moscow, Russia.

PETER V. ALEKSEEV

Institute of Global Economy and International Finance, Financial University under the Government of Russian Federation, Moscow, Russia.

ALEXANDER L. CHUPIN

Law Institute, Peoples' Friendship University of Russia (RUDN University), Moscow, Russia.

Abstract

Mobile money transfer systems (MMTS) in countries with limited banking are increasingly becoming the mainstream banking system. The analysis of transactions performed on these systems helps to detect fraudulent and criminal activities. This paper introduces a novel visual analytics framework for visual analysis and exploration of mobile money transactions. Our system enables the empirical analysis of mobile money transactions data using multiple views to reveal the temporal, geospatial, and categorical aspects of the transactions. Several challenges were identified related to the given MMT datasets through the process of implementing of fraud detection framework. In addition, as a step towards recognizing the difficult task of developing versatile and flexible fraud detection models, this work proposes concepts to address the identified challenges.

Keywords: Visual analytics, Fraud detection, Financial data visualization, Mobile money.

1. Introduction

Mobile money transactions systems are intensively used in countries with limited banking. For any monetary system, it is necessary to analyze transactions to detect criminal and fraudulent activities such as money laundering and terrorism financing. The current traditional methods cannot deal with multi-feature datasets like mobile transactions datasets; to deal with this task some new sophisticated methods are needed. In an attempt to solve this problem, we present MtVis a visual analytics system for mobile transactions analysis and exploration. We used a binary neural network classifier to perform fraud detection. During the implementation of our model, we identified several challenges specific to the given MMT datasets. This work offers the principles used to overcome these challenges as a step towards understanding the difficult task of designing flexible and scalable fraud detection models. In this article, we provide an overview of MtVis along with a concise analysis of how to develop an effective fraud detection model for mobile money payment services generated datasets. In the end, we provide an analysis of the framework by two domain experts.

The rest of the paper is organized as follows: Section 2 briefly explains the related work; Section 3 discusses fraud detection models; Section 4 presents the Visual Components of Mtvis; Section 5 presents the case studies for framework analysis; Section 6 describes the implementation of our system, we finally conclude the significance and importance of our framework in section 7.

2. Related Works

For financial data, a range of research studies on visual analytics have been carried out. A large number of tools have been designed to assist analysts in carrying out efficient risk management and decision-making in the enterprise. Our tool for mobile money transactions analysis is informed from the fields of financial data visualization and financial fraud detection.

2.1. Financial data analysis and visualization

A large number of visualization tools have been developed with growing concerns of organizations to turn their information into valuable assets. A survey of these systems is presented by Sungahn et al. [1] while Chang et al. [2] presented wire vis a financial time-varying data visual analytics. Didimo et al. [3] introduced VisFan, a network visualization system for financial data. Netsuite [4] is a financial planning visual analytics tool for decision making. Mobile money services, however, are subsystems of the traditional banking system with their own features and obstacles. The two major operators of mobile money services are Mpesa [5] and Orange money [6] deployed in Africa and the Middle East regions. Works specific to MMTS are a few due to its relative newness. Novikova et al. [7] worked on anomalous activity visualization in MMTS, and Gaber et al. [8] studied the behavior of users in an MTTTS environment.

2.2. Financial fraud detection

Fraud is the source of significant losses in financial systems. Various models of fraud detection have been proposed to resolve this problem. Abbasi et al. [9] designed an excellent framework for detecting financial fraud using meta-learning. Adedoyin et al. [10] used case-based reasoning to predict fraud in mobile money transfers. Kappelin and Rudvall [11] used statistical methods to tackle the problem of detecting fraud in Mobile Money Transfer systems. Novikova et al. [7] used the radviz [16] visualization technique to detect anomalous activities in the MMT environment while Didimo et al. [3] used network visualization to detect financial crime. Albashrawi et al. [12] did a review on fraud detection techniques using data mining. To provide researchers with experimental data, Lopez-Rojas et al. [13] designed Paysim which is a financial mobile money simulator for detecting financial fraud, this simulator can be used to simulate mobile transactions and generate data similar to the original dataset.

Current approaches are either based on the consumer or the financial statement as an entity, while our system uses transactions as an entity. The analysis of transactions as an entity helps to understand the overall state of MMT systems and reveals valuable trends useful in the fight against financial crime and terrorism financing.

3. Proposed Model for Fraud Detection

3.1. Dataset description

Our dataset was collected from 12 months of activity from a mobile network operator, it contains more than 400000 transactions. For security and privacy issues, the original dataset was processed to remove the sensible information. The basic entity of our data is a transaction; each transaction record is characterized by the following features.

- Type: The type of the transaction performed, taken as value: Cash-in, Cash-out, Transfer, Payment, Debit
- Location: Localization of the transaction.

- Amount: Amount of the transaction
- NameOrigin: Name of the sender
- NameDest: Name of the receiver
- BeforeBalanceSender: Balance of the sender before the transaction
- AfterBalanceSender: Balance of the sender after the transaction
- BeforeBalanceReceiver: Balance of the receiver before the transaction
- AfterBalanceReceiver: Balance of the receiver after the transaction
- Time: Time of the transaction

Our dataset also contains bench-marked transactions reported to be fraudulent, flagged by fraud or non-fraud features (Table 1). However due to the scarcity of bench-marked mobile transactions datasets and their private nature we used PaySim [13] to generate mobile transactions datasets based on our original dataset without changing our original features.

Table 1: Sample of data generated using paysim

Type	Amount	NameOrig	OldBalanceOrig	NewBalanceOrig	NameDest	OldBalanceDest	NewBalanceDest	IsFraud
CASHOUT	134991.97	C576685194	0	0	C36322011	391909.49	526901.46	0
TRANSFER	37991.39	C26423367	13136	0	C1764084529	0	37991.39	0
TRANSFER	3845765.36	C196788126	3845765.36	0	C1000407130	0	0	1
PAYMENT	16871.11	C652001878	0	0	M849216230	0	0	0
CASHIN	252508.29	C736570021	1828637.42	2081145.71	C928094130	27284930.5	27032422.21	0

3.2. Challenges in designing the proposed fraud detection model

Fraud detection is a classification problem, given $X_1...n$ features we are trying to output a $y_1...n$ classes, but some algorithms are more appropriate than others to perform this task because of their characteristics and the fact that the characteristics of fraud can differ over time. It requires various data preparation to obtain good results. In this section, we briefly describe the challenges in designing a good model for fraud detections in MMTS and elaborate on how to overcome these challenges.

- **Data Scarcity:** The lack of datasets on mobile money [18] transactions to perform research in the domain of fraud detection is a big problem. To overcome this, a simulator such as paysim [10] can be used to generate transactions data from a given input of the original dataset.
- **Imbalanced Data:** The big challenge [17] in modeling a fraud detection model as a classification problem is that the majority of real-world transactions are not fraudulent.

There are different ways of dealing with imbalanced data:

- SMOTE [14]: Synthetic Minority Over-sampling Technique. It consists of creating new synthetic samples by interpolating new points between marginal outliers and inliers.
- Oversampling: A common technique to deal with imbalanced data is oversampling [16]. It consists of creating new observations in our data belonging to the under-represented class.
- Under Sampling: It is the opposite of oversampling, it consists of reducing the observations of the dominant class.
- Choosing the right model: Fraud detection is a classification problem [19], so various classifiers can be used to solve it. However, since fraud features can change over time, we found that some classifiers are more suitable than others. For example, using a decision tree classifier [20] and a naive Bayesian classifier [21] we obtained good results on our dataset after performing feature engineering. Compared to other models such as a binary neural classifier and Gated Recurrent Unit Neural Network, we found that these models are not good due to their reliance on engineered features. The neural classifiers are free and much more versatile to adapt to new characteristics of fraud. We, therefore, chose the Gated Recurrent Unit Neural Network to perform the fraud detection.

3.3. The proposed Gated Recurrent Unit Neural Network model

As explained in section 3.2, we used a GRU neural classifier [22] to implement the fraud detection module of our system. The model is characterized as follows:

- Model overview: Figure 1 shows the flowchart of the proposed fraud detection model. First, we performed the preprocessing operations like numericalization (text to number) and normalization on the original dataset. The SMOTE algorithm is then used to process the samples which have low frequency in the training datasets. Subsequently, the analyzed data is used to train the GRU network.

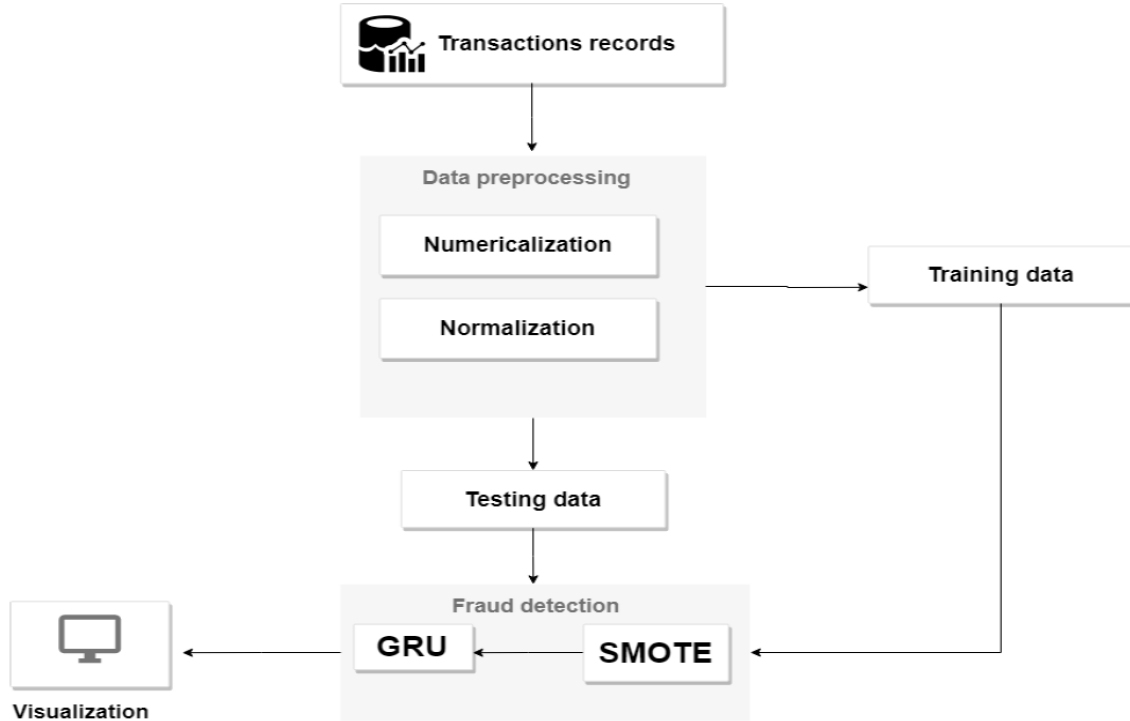


Figure 1: Fraud detection module

- Model parameters: The following parameters are used for the model: (a) Batch size and the number of neurons in all three (input, output, and hidden) layers of the GRU network during network training. (b) The parameters for Adam algorithm include step size, damping decrements for first-order and second-order exponential and nonzero constant. (c) Oversampling rate and the number of nearest neighbors for the SMOTE algorithm. The specified values for the parameters used in the experiments are shown in Table 2.

Table 2: Models parameters

Algorithm	Parameter	Value
SMOTE	Nearest neighbors	65
	Over-sampling rate	600%
GRU	Input layer nodes	122
	Hidden layer neurons	75
	Output layer neurons	5
	Batch size	500
Adam	Step size	0.001
	Damping Decrements for First-order exponential	0.9
	Damping Decrements fo Second-order exponential	0.999
	Non-zero constant	10^{-8}

- Metrics: To measure the effectiveness of the model, the indicators for evaluation which include ACC (accuracy), DR (detection rate), and FDR (false detection rate) were used to evaluate the performance of the proposed model for detection and then compare it with other state-of-the-art intrusion detection models and methods. The following equations are used to obtain these performance indicators:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$DR = \frac{Tp}{FN + TN} \quad (2)$$

$$FDR = \frac{FP}{TN + FP} \quad (3)$$

Where TP indicates the true positive, which means normal samples are predicted correctly, TN indicates the true negative, which means fraudulent samples are correctly

detected, FP represents the false positive which means fraudulent samples are mistakenly classified as normal and FN represents false negative which means normal samples are wrongly predicted fraudulent.

Experiment and results: To get optimal results for the fraud detection module multiple experiments were performed and parameter tuning was done such as the number of neurons in the hidden layers for the GRU neural network and oversampling rate and number of nearest neighbors for the SMOTE algorithm. Optimal values were chosen empirically for each parameter. The optimal accuracy was (99.33% ACC, 99.25 DR , 0.093% FDR) obtained with a system architecture of [122,200,1].

4. Visual Components

After performing the classification on input data using the GRU Neural Network, we feed the visualization components with the classified data for visualization and analysis. The system is composed of three main views each one performing a specific task: Fraud analysis view, Geospatial Analysis view, and Temporal analysis view. In this section, we describe each view and explain its functionality.

4.1. Fraud analysis View

The fraud analysis view (Figure 2) is our system's primary view and allows the user to understand the key payment patterns.



Figure 2: Fraud analysis view displaying (a) the bubble chart with clusters of transactions,(b) a chord to show intercity relations, (c) a bar chart displaying the state of the customers balance before and after performing the transactions (d) a parallel coordinates graph plotting multiple features for each city

It is composed of the following components:

- a) **Bubble chart:** The bubble chart displays clusters of transactions according to their type and their fraudulent or non-fraudulent nature. The first level is a green cluster containing the normal transactions and a red one containing the fraudulent transactions. In the second level, the transactions are organized by category using color encoding and in the third level each transaction is displayed as a circle, the size of the circle is proportional to the amount of money involved in the transaction. Each level of the bubble chart is zoomable and interactive which provides the user with a better analytic experience.
- b) **Chord chart:** The chord chart shows the relationship between the locations of the transactions. This chart is useful to know the way of interactions among users from different cities and to assess the volume of the transactions between cities.
- c) **Bar chart:** The filterable bar chart displays the situation of the customers' balance before and after performing the transaction. The variation of the balance is an important fraud feature. A query panel allows the user to query different transactions and is used to make hypotheses and check them against the dataset.
- d) **Parallel coordinates graph:** The parallel coordinates graph is a multi-functional representation of transactions for each city. The parallel coordinates graph aims to allow the user to easily analyze the various transaction dimensions per city.

4.2. Geospatial Analysis View

The geospatial analysis view allows the user to grasp the geographical state of the transactions. It is composed of the following three layers. A scatter in Figure 3 (a) shows the geographical distribution of fraudulent transactions. The radius of the scatter is proportional to the number of fraudulent transactions. A hexagon layer in Figure 3 (b) is used to display the volume of transactions performed per city. An arc layer in Figure 3 (c) displays the transaction flow among cities. Using the thickness of the arcs and colors we abstract the volume of transactions In and Out between two cities.

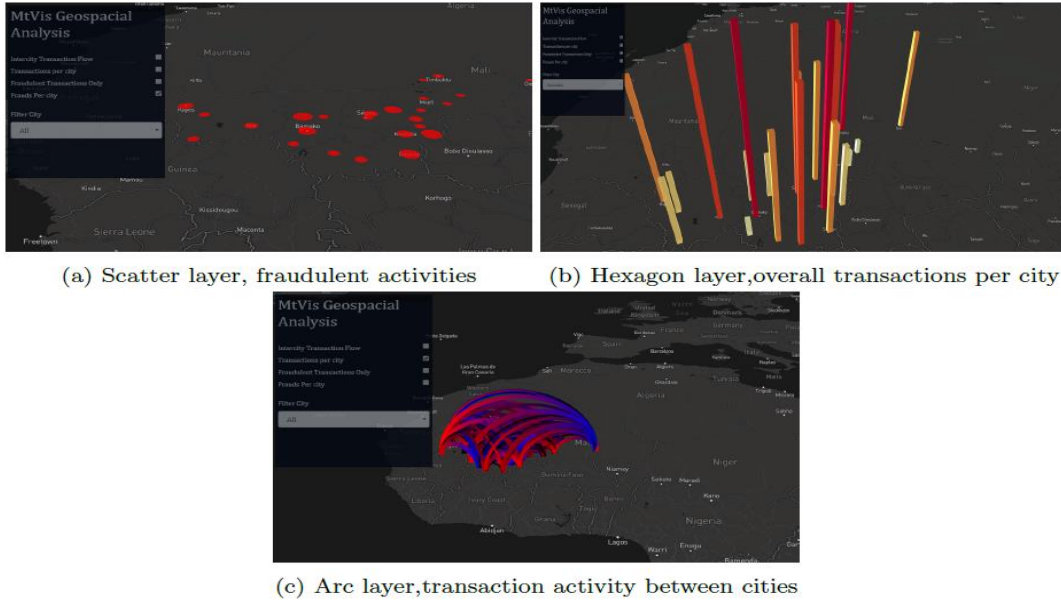


Figure 3: Layers of the geospatial analysis view

4.3. Temporal Analysis View

The temporal analysis view (Figure 4) is used to display the time aspect of the transactions and is composed of three charts:

- A calendar activity view shows the distribution of transactions per day.
- A radar view is used to display the variations of activity at different times of the day. This is useful to identify the exact time when fraudulent transactions occur the most.
- At the bottom there is an interactive and filterable table displaying the classified data, allowing the user to deeply analyze it.

To enhance the analysis and exploration experience, we implemented a wide range of interactions and filters. The user can hover on components for more details and filter the information.



Figure 4: Temporal analysis view, (A) a calendar activity view displaying intensity of transactions per day, (B) a radar chart displaying the time distribution of transactions per category and (C) a table view displaying the raw data.

5. Case studies and Expert Evaluation

We performed two case studies with two experts (hereinafter referred to as Expert A and Expert B) to verify the usefulness of our system. Each expert used our system to analyze the transactions and gave us feedback.

5.1. Case study 1: Terrorism financing and money laundering

Expert A is an anti-terrorism agent who used the system (Figure 5) to monitor transactions between major cities and sensitive areas where terrorists were based using the fraud analysis view filtering method. The expert then analyzed the transactions that took place between sensitive areas using the feature of geospatial analysis. Future investigations will be carried out into the fraudulent transactions found. He then commented on the system. "The geospatial view is really important when monitoring transactions for terrorist financing and analyzing transactions in sensitive regions." Money laundering happens when illegal money is introduced into the financial system through placements. To track such placements our transaction query can be used to filter transactions in which big amounts of money are placed (cash in).

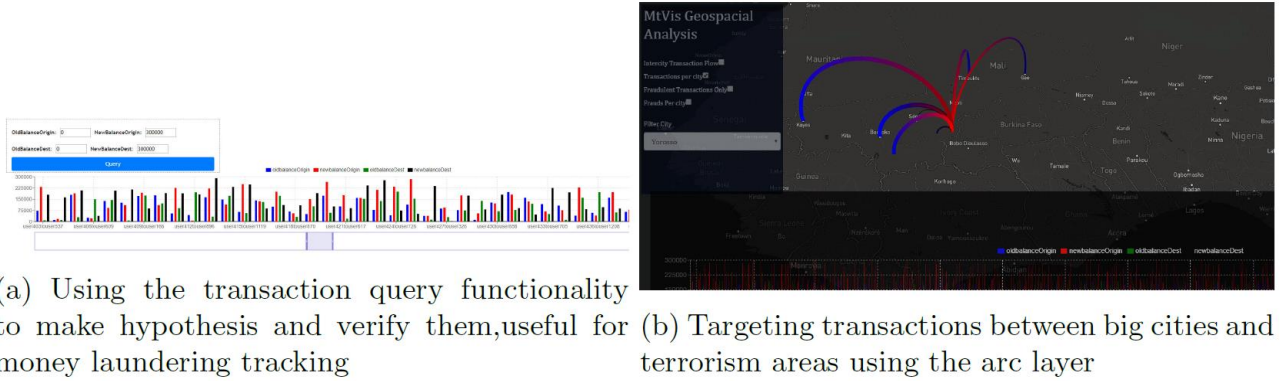


Figure 5: Case study 1

5.2. Case study 2: Users' behavior

Expert B is a senior financial advisor for mobile money who used the system (Figure 6) to analyze the behavior of users in different cities. He found that users in the biggest cities tend to perform transactions involving cash in, cash out, and payment to vendors, the users in small cities on the other hand tend to receive money, and they perform more cash out, transfer, and payment activities. Using the temporal analysis view the expert analyzed the temporal behaviors of their customers and found interesting trends in the timing of the transactions. The expert stated, “This is very useful to get relevant information when deploying our new services and planning to cover new regions”.



Figure 6: Case study 2

6. System Implementation

Our system is a client-server application with a ReactJS [23] front and a Flask [24] back end. The Fraud detection module was developed using the open-source deep learning framework PyTorch [25-31]. The Detection module is deployed on the server where the input data is preprocessed and sent to the front end for visualization.

7. Conclusion

We presented MtVis as a visual analytics framework for analyzing and exploring mobile money transactions and have shown the significance of our tool in the exploration and visual analysis of a mobile transaction system. Our system has significantly been efficient in detecting the planning of fraud. We also demonstrated the importance of the transaction as an entity in a mobile money transfer system in detecting anomalous activities. However, in this work, we only focused on the transaction entity which is a small component of a mobile money transfer system. In future works, we will dive deeper into other components such as user behavior and user categorization.

Acknowledgment/Funding: Alexander L. Chupin has been supported by the RUDN University Strategic Academic Leadership Program.

REFERENCES

- [1] Ko S, Cho I, Afzal S, Yau C, Chae J et al. A survey on visual analysis approaches for financial data. In: Computer Graphics Forum; Groningen, Netherlands; 2006. pp. 599-617. doi: 10.1111/cgf.12931
- [2] Chang R, Ghoniem M, Kosara R, Ribarsky W, Yang J et al. Wirevis: Visualization of categorical, time-varying data from financial transactions. In: IEEE Symposium on Visual Analytics Science and Technology; Sacramento, CA, USA; 2007. pp. 155-162. doi: 10.1109/VAST.2007.4389009
- [3] Didimo W, Liotta G, Montecchiani F, Palladino P. An advanced network visualization system for financial crime detection. In: IEEE pacific visualization symposium; Hong Kong, China; 2011. pp. 203-210. doi:10.1109/pacificvis.2011.5742391
- [4] NETSUITE(2019). Business Management Software [online].Website <https://www.netsuite.com> [accessed 20 01 2022].
- [5] Mpesa(2019). Vodafone mPesa India [Online]. Website <https://www.mpesa.in> [accessed 20 01 2022].
- [6] Orangemoney (2019). Transfert d'argent s ecuris e de mobile `a mobil [Online]. Website <https://orangemoney.orange.fr> [accessed 20 01 2022].
- [7] Novikova E, Kotenko I. Visual Analytics for Detecting Anomalous Activity in Mobile Money Transfer Services. In: International Conference on Availability, Reliability, and Security; Fribourg, Switzerland; 2014. pp. 63-78. doi:10.1007/978-3-319-10975-6
- [8] Gaber C, Giot R, Achemlal M, Hemery B, Pasquet M et al. Analyse des comportements dans un syst`eme de transfert d'argent sur mobile. In: 8`eme Conf`erence sur la S`ecurit`e des Architectures R`eseaux et Syst`emes d' Information(SAR SSI); Mont-de-Marsan, France; 2013. pp.10.(in French)
- [9] Abbasi A, Albrecht C, Vance A, Hansen J. Metafraud: A meta-learning framework for detecting financial fraud. *Mis Quarterly* 2012; 36(4): 1293-1327. doi:10.2307/41703508
- [10] Adedoyin A, Kapetanakis S, Samakovitis G, and Petridis M. Predicting fraud in mobile1 money transfer using case-based reasoning. In: International Conference on Innovative Techniques and Applications of Artificial Intelligence; Springer, Cham, Switzerland; 2017. pp 325–337. doi:10.1007/978-3-319-71078-5 28

- [11] Kappelin F and Rudvall J. Fraud detection within mobile money: A mathematical statistics approach. Master, Blekinge Institute of Technology, Karlskrona, Sweden, 2015.
- [12] Albashrawi M, Lowell M. Detecting financial fraud using data mining techniques: a decade review from 2004 to 2015. *Journal of Data Science* 2016; 14(3): 553-570.
- [13] Lopez-Rojas E, Elmir A, Axelsson S. PaySim: A financial mobile money simulator for fraud detection. In: 28th European Modeling and Simulation Symposium; Larnaca, Cyprus; 2016. pp.249-255.
- [14] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 2002; 16: 321-357. doi:10.1613/jair.953
- [15] Jack W, Ray A, Suri T. Transaction networks: Evidence from mobile money in Kenya. *American Economic Review* 2013; 103(3): 356-361. doi:10.1257/aer.103.3.356
- [16] Sharko J, Grinstein G, Marx KA. Vectorized radviz and its application to multiple cluster datasets. *IEEE transactions on Visualization and Computer Graphics* 2008; 14(6): 1444-27. doi:10.1109/tvcg.2008.173
- [17] Sun Y, Kamel MS, Wong AK, Wang Y. Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition* 2007; 40(12): 3358-3378. doi:10.1016/j.patcog.2007.04.009
- [18] Chu J, Efendiev Y, Ginting V, Hou TY. Flow based oversampling technique for multiscale finite element methods. *Advances in Water Resources* 2008; 31(4): 599-608. doi:10.1016/j.advwatres.2007.11.005
- [19] Major JA, Riedinger DR. EFD: A hybrid knowledge/statistical-based system for the detection of fraud. *Journal of Risk and Insurance* 2002; 69(3):309-324. doi:10.1111/1539-6975.00025
- [20] Breiman L. *Classification and Regression Trees*. NY, USA: Routledge, 2017. doi: 10.1201/9781315139470
- [21] Corani G, Benavoli A, Demšar J, Mangili F, Zaffalon M. Statistical comparison of classifiers through Bayesian hierarchical modelling. *Machine Learning* 2017; 106(11): 1817-1837. doi:10.1007/s10994-017-5641-9
- [22] Jozefowicz R, Zaremba W, Sutskever I. An empirical exploration of recurrent network architectures. In: *International Conference on Machine Learning*; Lille, France; 2015. pp. 2342-2350.
- [23] Aggarwal S. *Modern Web-Development using ReactJS*. *International Journal of Recent Research Aspects* 2018; 5(1): 133-137.
- [24] Grinberg M. *Flask web development: developing web applications with python*. Sebastopol, CA, USA: O'Reilly Media, Inc., 2018.
- [25] Ju S, Ramjee S, Yang D, El Gamal A. A PyTorch Framework for Automatic Modulation Classification using Deep Neural Networks. *The Summer Undergraduate Research Fellowship (SURF) Symposium 2018*; Paper 77.
- [26] Shaikh, Zaffar Ahmed, Abdullah Ayub Khan, Laura Baitenova, Gulmira Zambinova, Natalia Yegina, Natalia Ivogina, Asif Ali Laghari, and Sergey Evgenievich Barykin. "Blockchain Hyperledger with Non-Linear Machine Learning: A Novel and Secure Educational Accreditation Registration and Distributed Ledger Preservation Architecture." *Applied Sciences* 12, no. 5 (2022): 2534.
- [27] Shaikh, Z. A., & Khoja, S. A. (2013). Higher education in Pakistan: An ICT integration viewpoint. *International Journal of Computer Theory and Engineering*, 5(3), 410.
- [28] Shaikh, Z. A., & Lashari, I. A. (2017). Blockchain technology: The new internet. *International Journal of Management Sciences and Business Research*, 6(4), 167-177.
- [29] Shestak, V., Gura, D., Khudyakova, N., Shaikh, Z. A., & Bokov, Y. (2020). Chatbot design issues: building intelligence with the Cartesian paradigm. *Evolutionary Intelligence*, 1-9.
- [30] Shaikh, Z. A., Umrani, A. I., Jumani, A. K., & Laghari, A. A. (2019). Technology enhanced learning: a digital timeline learning system for higher educational institutes. *International Journal of Computer Science and Network Security*, 19(10), 1-5.
- [31] Shaikh, Z. A. (2018). Keyword Detection Techniques. *Engineering, Technology & Applied Science Research*, 8(1), 2590-2594.