ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 12, Issue 4, 2018

ENHANCING FINANCIAL FRAUD DETECTION USING DEEP LEARNING TECHNIQUES WITH SEQUENTIAL

[1]Karthik Kushala
Celer Systems Inc,Folsom,California,USA
karthik.kushala@gmail.com

S. Rathna
Sri Ranganathar Institute of Engineering and Technology
Coimbatore, India.
rathnajack@gmail.com

ABSTRACT

A Financial fraud detection deep learning model involves a sequenced attention-based Long Short-Term Memory (LSTM). This attention mechanism very sensitively tunes itself to minor signals in collecting fraud, and the other dimension-reduction method is Principal Component Analysis (PCA). The zeroth LSTM in the model classifies time-sequenced transactional data as either fraudulent or non-fraudulent. So, it can detect the fraud in real-time within the financial set-up because this model scored very good marks regarding accuracy, precision, sensitivity, and specificity. This work, therefore, shows that LSTMs, in combination with their attention mechanisms with deep learning techniques, will significantly promise the future for fraud detection systems. The performance of the model is tested based on a confusion matrix, with an accuracy of 0.9885, a precision of 0.967, a sensitivity of 0.975, a specificity of 0.971, and 0.989 in F-measure. The Negative Predictive Value (NPV) is 0.986, which is indicative of superior performance in identifying true fraudulent transactions. This methodology is very effective in integrating structure learning and temporal pattern discovery for detecting financial fraud.

Keywords: Principal Component Analysis (PCA), Fraud Detection, Long Short-Term Memory (LSTM), Fully connected layer.

1. INTRODUCTION

The financial fraud detection is undergoing growing importance and rising nuisance, with the trend of fraud changing well over time. Largely traditional detection methodologies could not do well because the amounts and complexities of transactional data are huge [1]. The invention of deep learning is said to have opened new windows in the market for fraud detection relative to accuracy and enhancement. Moreover, the understanding is that deep-learning systems'- especially sequential models' prowess- could very well capture highly complex patterns with dependency on time [2]. The paper talks about a sequential attention-based long-short-term memory model competent in handling time-sequenced financial transaction data [3]. The meant attention mechanism refines a further capability of the model to attend to local areas in the data that are beneficial hollow of anomaly detection and that would otherwise be close to impossible [4]. Dimension reduction through PCA and parameter relevance computation is all part of it. The whole exercise has been to further enhance the operational speed and performance of the model. Thus, the methodology developed intends to form a strong, large fraud detection system that could classify transactions into fraud or legitimacy [5]. Thus, this work is an endeavor to establish a balance in the new fraud detection in finance between the deep learning technologies and the sequential architecture that will be an excellent fit for real-time systems for fraud detection and prevention in financial transactions . The present paper discusses a model for financial-fraud detection that employs state-of-the-art methods for performing the modeling and, hence, performance and accuracy enhancement. This model considers raw data pre-processing and cleaning processes for handling missing values that could be used for any further purposes in practical analysis [6]. A core model is proposed for a Long Short-term Memory Network with an attention mechanism. Since an LSTM is great at modeling sequential data such as financial transactions, the model could preserve long-term patterns and dependencies through time. Attention thus builds on the powers of LSTM by allowing the model to focus on the important components in its input sequence to uncover the very subtle traces of fraud [7]. For diversified production, a group of bidders submits proposals that will enable them to bid on the various alternative combinations and substitutes that can be then put together. Hence, the output of the sigmoid function is a probability, varying from 0 to 1, where 1 indicates fraud and 0 indicates no fraud [8]. Hence, employing all possible techniques could provide one with a reliable, accurate, and scalable system for fraud detection. Nevertheless, there are limitations in the system proposed for financial fraud detection. Most are exclusively usable in LSTM networks [9]. This is simply because their ability to capture temporal dependencies is better,

however, they become heavy on computations and time-consuming for training before they can show any efficiency. The other is the Attention Mechanism. This one has very high powers and thus makes all things a bit messy, so it becomes a little less interpretable [10]. Of course, this is one very important factor in fraud detection scenarios. Most likely, the principal component analysis (PCA) step provides dimensionality reduction that leaves out many probable important features unless well tuned, which is going to affect the model performance more. Besides that, the model depends on the quality and completeness of transaction data to be labeled according to what they are supposed to be, which does not apply to real situations [11]. This can also cause a bias that affects predictions from here, the outcome is likely influenced. Finally, it can also be argued that the model depends on a single dataset, which in turn limits model transferability to the various financial systems or environments. It ensures that there are certain potential fraud patterns not included in the capture. Other improvement areas include scalability, explainability, and adaptation to many more forms of evolving fraud schemes. The contribution of the paper is below;

- LSTM and Attention Integration: The paper introduces an LSTM model with an attention mechanism intended to bolster the effective detection of frauds via focusing on relevant areas for time-sequenced data.
- Dimensionality Reduction with PCA: Whereas PCA reduces the complexities of the dataset, making it easier to handle data, in essence, it is for the sake of faster computational times for real-time fraud detection.
- Real-Time Fraud Detection System: This paper talks about developing and implementing a large-scale and accurate system that can classify any transaction as either fraud or legitimate in real-time.

2. LITERATURE SURVEY

The financial institutions will experience immeasurable challenges as a result of the rapidly proliferating online threats. Hence, the detection and prevention functionalities have utmost importance. Progress in the application of machine learning and data mining for modeling fraud detection in finding unknown patterns in massive data processing that always works with some degree of success is scant. The paper proposes to develop a hybrid fraudulent detection system that activates two deep learning models: an autoencoder and an LSTM RNN, so that the system is most efficient in detecting early signs of financial fraud [12]. It uses oversampling techniques to counter the deficiencies caused by imbalanced datasets-particularly a relevant application problem in the field of fraud detection. Hybrid approaches are known to catch more fraud incidents as compared to mathematical methods of machine learning. Good recall as well as precision values would be achieved through experimental evaluation to take forward fraud detection systems. Even though the mode of e-commerce is changing swiftly, it allows every customer to pick items at the best price, quality, and quantity without hassles [13]. Although this convenience has provided opportunities to fraudsters as well, the most prevalent form of transaction has become that of credit card payment. To bring automated anti-theft mechanisms from fraud, banks should put up a fraud detection ensemble based on deep recurrent neural networks and artificial neural networks. Not only for that, but also, the training scheme for the voting mechanism in this model is something very new. Results from experiments show that the proposed model is indeed superior to existing models, yet more efficient concerning real-time performance-that makes it practical for the implementation of fraud detection systems [14]. These deep-learning algorithms are very much like their other counterparts in the field of fraud detection-only that this time they will learn the whole-of-the-entity-behavior sequences and not just snapshot data. Define a new and innovative paradigm along the lines of modeling the behavior sequences that are driven through events that allow them to fit into truly sequential patterns. The self-history attention mechanisms and personalized forget gates create long-term dependencies for the sake of foreground novelty, whereas background source knowledge feeds into the interaction module for improving performance. The real-time testing runs using a telecommunications dataset established the detection of fraud while differentiating variations in terms of an activity performed fraudulently from the regular ones [15]. The text describes a new composite architecture for Deep Sequential Learning DSL models called the "contained-in-between CIB model," with the expected maximization EM model, DSL model, and upper EM-classifier as subcomponents. This structure can truly take a front-row seat in the most commanding application fields of distributed fraud detection systems, where data is generally characterized by complex interrelations all around [16]. The model shows exceptionally good performance with transaction fraud detection and probably represents a place between deep learning and structured model ensembles.

3. PROBLEM STATEMENT

The study, therefore, focuses on developing efficient and effective financial fraud detection systems for the possible detection of fraud in financial transaction data [17]. Most of these traditional methods of fraud detection mask the complexity of the data, especially when they are large, high-dimensional datasets, and delve into the temporal patterning of financial transactions [18]. A further step of intervention in the approaches has

involved deep learning using a sequential attention-centered LSTM model and Attention Mechanism to address these issues. It creates sequential dependencies and critical transaction modalities by keeping a tab on credible previous transactions, thus improving accuracy in fraud detection, since this architecture is oriented towards time series. The architecture employs PCA for dimensionality reduction and uses an LSTM with attention to mention the truly relevant features and temporal relations [19]. It is geared to developing a strong fraud detection system to classify transactions efficiently as fraudulent or non-fraudulent, thus contributing to a more realistic and almost real-time, scalable solution in the detection of fraud within financial transactions [20].

4. PROPOSED METHODOLOGY

This proposed financial fraud detection model consists of a sequential attention-based LSTM and an Attention Mechanism model that takes the Financial Fraud Detection Dataset as a fulcrum for the entire methodology process. After data collection is data pre-processing, where the entire dataset is cleansed and missing values. That way, the data is prepped for the next step. Next, Principal Component Analysis (PCA) is done for feature extraction. PCA is highly effective in compressing the dataset dimensionality by eliminating unimportant forensic features, thereby facilitating more computational efficiency in the modeling process. The sequential architecture of the model now merges the attention mechanism with the LSTM network as one module. The LSTM works to maintain a record of ongoing occurrences, being most appropriate for sequential data such as financial transactions that are usually time-stamped and ordered. This attention mechanism proposes to boost LSTM's capabilities since localized attention enables the model to consider the most prominent parts of a sequence, which may significantly benefit its ability to identify unusual behaviors with transaction records. At the very least, this will commit the result to a fully connected layer with the last component being a sigmoid activation function. The function classifies any input-case data as either financial fraud detected or no financial fraud detected. The output from the sigmoid function is a value that ranges from 0-1, which represents the probability of fraud, whereby 1 means fraud and 0 means no fraud. This proposed model represents an advanced manner whereby the detection of financial fraud utilizes the methodology of deep learning in all of its modernity. Hence, the financial fraud detection system is made more precise and reliable for the detection of fraudulent activity in transactions.

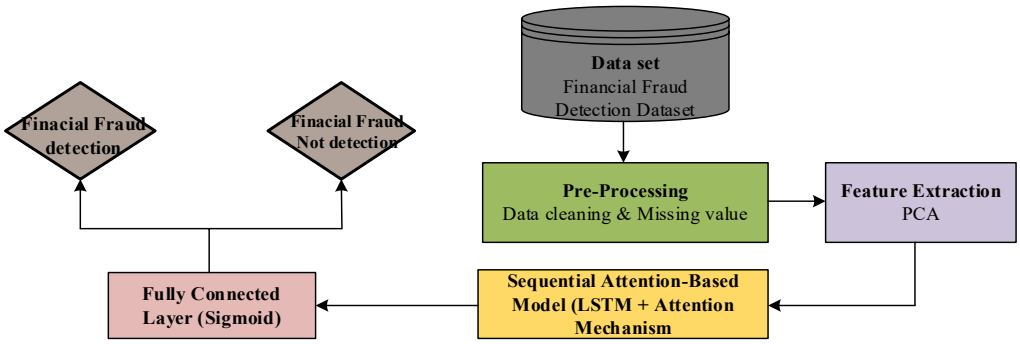

Figure 1: Architecture diagram

4.1 DATA PRE-PROCESSING

Thus, the data pre-processing takes charge of preparing raw data for further processing and analysis or building a model on it. In this phase in transform data is transformed into a form that allows for feature extraction and training on models. Such pre-processing may bring data to value ranges (normalization/standardization), category data encoding, sum, and aggregate data sets. The sole objective of data pre-processing is to have all the data presented consistently across feature extraction for superior training performance of a model.

4.2 DATA CLEANING

Data cleaning means resolving inconsistencies, errors, and any irrelevant items that the dataset may possess. Those duplicate records are distorting the analysis. Remove errors due to misformatting and fix incorrect values that would cause inconsistency among its features. The data must be cleaned and formatted accordingly, since it influences the quality of the analyses and the overall performance of the model.

4.2 HANDLING MISSING VALUES

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 12, Issue 4, 2018

Missing data is one of the critical areas of data cleaning, as any absence in data may bias the analysis and findings of the subsequent model. There are many possible ways to treat the different types of missing values. First and probably simplest of them all is imputation, which essentially uses some measure of the target-feature value, mean, median, or mode-as a replacement for the missing data. For cases where this approach is too simple, predictive modelling techniques and neighbour-based methods can ascertain missing values by seeing how they relate to the other values in the dataset. If the percentage of values missing is so high that it would be unreasonable to impute, the following choice would be to drop certain rows or columns of data to further prune it for actual analysis and modelling.

a) Mean/Median Imputation

Imputation of missing values usually involves the statistical mean or median. Mean imputation fills missing values with the mean of the observed values in that feature, whereas median imputation replaces them with median values calculated after ordering. Median is preferred over mean usually when there are outliers, since it is not strongly affected by extreme values, thus making a more robust imputation method in terms of skewed distributions. They maintain the coherence of the dataset at the expense of a very small amount of information.

4.3 PRINCIPAL COMPONENT ANALYSIS(PCA)

Principal Component Analysis(PCA) is a statistical method to reduce the dimensionality of a data set and extract some important features present in the data. The principal component uncorrelated variable represents a small proportion of several variables, so that most of the variance (or information) is retained. PCA can theoretically be represented in a flowchart or architecture diagram concerning a data pipeline. Since it usually partakes in either pre-processing or the feature extraction part of machine learning workflows, it can be put into an attractive diagram where it parameterizes the flow from high-dimensional data to the low-dimensional components to ease further analysis or model building.

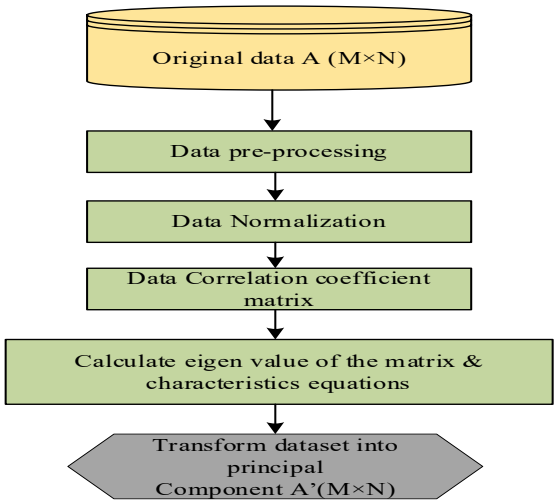

Figure 2: Flow chart of PCA

4.4 SEQUENTIAL ATTENTION-BASED MODEL

A sequential attention-based model is predominantly used to process the data by combining the input provided by Long Short-Term Memory (LSTM) and an Attention Mechanism. LSTM is an ideal network to manage a flow of events properly over time. In other words, it processes the sequential data, such as time-series or transaction-based data, which is very common for financial fraud detection. LSTM makes retention of all the important patterns and dependencies, usually over time, hence it can well capture the temporal relationship. The introduction of this attention mechanism to the LSTM would allow the model to focus most attention on what matters, thus allowing the model to behave in a way that the LSTM would from other inputs but across many of the actual inputs. Here lies the benefit of different parts of a sequence and assigning different weights to different parts of the sequence according to importance. The model thus would prioritize the most meaningful features, resulting in better detection of even minor pattern changes that might suggest illegalities. Overall, this LSTM and Attention Mechanism combined architecture is powerful in processing sequences of data and learning about failure modes concerning time dependencies and critical patterns within the data. This combination simply indicates that the model is capable of detecting even the very complex fraud patterns developing over time, thus improving the performance of all systems deployed for financial fraud detection. The mathematical formula of

the output of a sequential attention-based model combining an LSTM and an attention mechanism can be expressed as:

$$Output = \sigma(W_0 \text{ Attention(LSTM(X))}+b_0) \quad (1)$$

Where $\sigma$is the sigmoid activation function, W_0 is the weight matrix for the output layer, b_0 *is the bias term for the output layer,* Attention represents the attention mechanism, and LSTM(X) is the output of the LSTM network when processing input data X.

4.5 FULLY CONNECTED LAYER (SIGMOID):

A fully connected layer is one of the most significant parts of a detection model, which comes right after the sequential attention-based model (LSTM + attention mechanism). The work of a fully connected layer is to take the output of previously used layers and transform all this information to get the final prediction. The whole neuron of one layer is completely connected to each neuron of the current layer. Thus, the fully connected layer inputs being in the sigmoid activation function yield values between 0 and 1, which indicates the probability of the financial transaction being fraudulent. This is mainly because the sigmoid function is supposed to classify the transaction as financial fraud detected (1) or as no financial fraud detected (0) using already learned features and patterns from the dataset. It is a vital decision point for the model, as it gives the required output of binary classification on the problem of fraud detection. The mathematical representation of the connected layer with a sigmoid activation function while performing binary classification is simply

$$y = \sigma(W.x + b) \quad (2)$$

Where Y is output, with σ as the activation function being a sigmoid, W referring to the weight matrix corresponding to the complete connected layer, x being input from the previous layer, whereas b is the bias value associated with the complete connected layer.

5. DATASET DESCRIPTION

The Financial Fraud Detection Dataset is usually a collection of transactional data for detecting fraudulent activities in financial systems. This dataset usually contains a wide variety of attributes about financial transactions like transaction ID, amount, time-stamp, payment method, merchant details, and customer data. The data normally contains both fraudulent and non-fraudulent transactions with the objective of training machine learning models to classify the transactions into either legitimate or fraudulent [21]. The dataset is, anyway, very imbalanced, with comparatively far fewer fraudulent transactions in contrast to those that are legitimate; thus, creating real issues in developing accurate models. The transactions might encompass several feature types: transaction type, customer demographic, purchase history, geographical location of the transaction, etc. The target variable in the dataset usually signifies whether the transaction is fraudulent (1) or not fraudulent (0). The dataset may also have temporal patterns in transaction behaviour during time correlated with fraud detection, as fraudulent activities are ever-growing and evolving. The Financial Fraud Detection Dataset is often pre-processed for handling missing data, dealing with class imbalance, and outliers to make it suitable for developing as well as evaluating strong fraud detection models. The dataset can be used as a very valuable source for the development of a system that can automatically identify and thwart fraudulent transactions in financial services.

Dataset Link: https://www.kaggle.com/datasets/sriharshaeedala/financial-fraud-detection-dataset

6. RESULT AND DISCUSSION

With a prediction accuracy score of around 0.9885, the Financial Fraud Detection Model predicts around 98.85% of the samples correctly. The precision here states that of those tagged as frauders, 96.7% were indeed actual defrauders, yielding a precision value of 0.967. The model sensitivity of 0.975 can explain that the model could catch about 97.5% of the fraud cases, while the specificity of 0.971 indicates that it could correctly identify 97.1% of the time non-fraud cases, therefore greatly reducing false positives. F-measure 0.989 implies the perfect balance between the precision and recall, which gives confidence in actual fraud detection. The NPV of 0.986 indicates that the model will most probably detect fraud and predict concretely 98.6% of the non-fraud cases. Overall, it is very strong. The model detects fraud like a pro, but does has relatively more false positives than false negatives as well. This metric proves that it is a capable model for the detection of financial fraud and a very strong candidate for implementation in real-time detections in the financial systems. The model is also ensured to have high accuracy and operational efficiency in the detection by the deep learning paradigm with LSTM and attention mechanisms. Thus, this huge practical strength renders the model quite suitable for safety engagements where high precision in fraud detection is a must.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 12, Issue 4, 2018

6.1 CONFUSION MATRIX

As shown in the confusion matrix above, it can be seen how much the model is performing well in identifying the fraud. Out of all occurrences of actual fraud, 362 were detected as fraudulent, and the remaining 76 incidents of non-fraud were classified incorrectly as fraud (False Positive). On the other hand, the test misclassified only 3 cases of actual fraud as non-fraud (False Negative). However, out of the 393, they could predict well that they were not fraud (True Negative). This, in turn, gives rise to a matrix describing the detecting ability of the model in terms of the fraud along with its challenges, against minimizing the false positive counts, with adjustments made to keep the false negative numbers low.

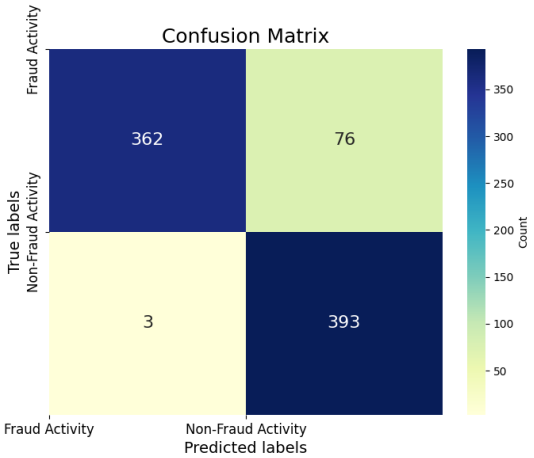

Figure 2: Confusion Matrix

6.2 PERFORMANCES MATRIX

The Financial Fraud Detection Model has an accuracy of 0.9885, meaning it classifies 98.85% of the samples correctly. It has a precision of 0.967, which means that 96.7% of the predicted fraud cases are typically actual fraud cases. It has a sensitivity of 0.975-97.5 of the time correctly identifies fraud cases-and a specificity of 0.971, where 97.1% of non-fraud cases are correctly identified. An F-measure of 0.989 and NPV of 0.986 illustrate that the model is performing significantly well to detect fraud with very few false positive detection rates.

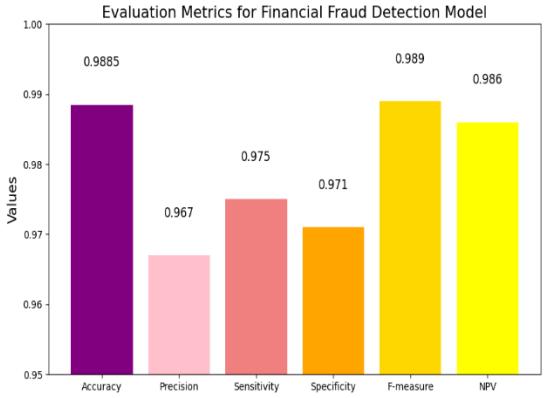

Figure 3: Performance matrix

7 CONCLUSION:

A proposed financial fraud detection model has made an efficient impact with a sequential attention-based LSTM architecture on time-distributed data to optimize the detection of fraudulent activities during financial transactions. This model is as high as 98.85% in accuracy performance and as high as 96.7% in precision, thus indicating a good model performance capable of correctly classifying cases into fraud and non-fraud ones. PCA has been the dimensionality reduction used here, while attention mechanism was paid to appropriate areas of transaction data, consequently improving the model's ability to detect more subtle patterns of fraud. So, it is for many other considerations, including dependency on the quality of labelled data and several other biases, that this

model presents a good potential for application in real life. However, it is a strong but efficient answer to financial fraud detection. Scalability requirements, along with the interpretability of models and adaptation to evolved fraud mechanisms, have to be proved to improve the model further.

References

[1] S. S. Parimi, "Leveraging Deep Learning for Anomaly Detection in SAP Financial Transactions," Nov. 17, 2017, *Social Science Research Network, Rochester, NY*: 4934907. doi: 10.2139/ssrn.4934907.
[2] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3784–3797, Aug. 2018, doi: 10.1109/TNNLS.2017.2736643.
[3] D. Abdelhamid, S. Khaoula, and O. Atika, "Automatic Bank Fraud Detection Using Support Vector Machines," 2014.
[4] R. Sarno, R. D. Dewandono, T. Ahmad, M. F. Naufal, and F. Sinaga, "Hybrid Association Rule Learning and Process Mining for Fraud Detection," 2015.
[5] Computer Department, SRESCOE Kopargaon, India, Ms. A. D. Pawar, Prof. P. N. Kalavadekar, and Ms. S. N. Tambe, "A Survey on Outlier Detection Techniques for Credit Card Fraud Detection," *IOSR J. Comput. Eng.*, vol. 16, no. 2, pp. 44–48, 2014, doi: 10.9790/0661-16264448.
[6] S. Haroon and D. K. Hussain, "Useful Data Mining Techniques to Prevent Online Auction Fraud Detection," vol. 2, 2015.
[7] E. Mbunge, R. Makuyana, N. Chirara, and A. Chingosho, "Fraud Detection in E-Transactions using Deep Neural Networks - A Case of Financial Institutions in Zimbabwe," vol. 6, no. 9, 2015.
[8] J. West, M. Bhattacharya, and R. Islam, "Intelligent Financial Fraud Detection Practices: An Investigation," in *International Conference on Security and Privacy in Communication Networks*, J. Tian, J. Jing, and M. Srivatsa, Eds., Cham: Springer International Publishing, 2015, pp. 186–203. doi: 10.1007/978-3-319-23802-9_16.
[9] Y. D. Hardas and A. Pawar, "A REVIEW OF VARIOUS CREDIT CARD FRAUD DETECTION TECHNIQUES," 2016.
[10] C. Liu, Y. Chan, S. H. Alam Kazmi, and H. Fu, "Financial Fraud Detection Model: Based on Random Forest," *Int. J. Econ. Finance*, vol. 7, no. 7, pp. 178–188, 2015.
[11] "Automatic ATM Fraud Detection as a Sequence-based Anomaly Detection Problem:," in *Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods*, ESEO, Angers, Loire Valley, France: SCITEPRESS - Science and and Technology Publications, 2014, pp. 759–764. doi: 10.5220/0004922307590764.
[12] P. JRana and J. Baria, "A Survey on Fraud Detection Techniques in Ecommerce," *Int. J. Comput. Appl.*, vol. 113, no. 14, pp. 5–7, Mar. 2015, doi: 10.5120/19892-1898.
[13] "International Journal of advanced studies in Computer Science and Engineering," vol. 4, no. 3, 2015.
[14] L. Nahar, I. Amir, and S. Shabnam, "A Comprehensive Survey of Fraud Detection Techniques," *Int. J. Appl. Inf. Syst.*, vol. 10, no. 2, pp. 26–32, Dec. 2015, doi: 10.5120/ijais2015451471.
[15] R. Saia and S. Carta, "Evaluating Credit Card Transactions in the Frequency Domain for a Proactive Fraud Detection Approach:," in *Proceedings of the 14th International Joint Conference on e-Business and Telecommunications*, Madrid, Spain: SCITEPRESS - Science and Technology Publications, 2017, pp. 335–342. doi: 10.5220/0006425803350342.
[16] A. Nahar, S. Roy, and S. Shabnam, "A Survey on Different Approaches used for Credit Card Fraud Detection," *Int. J. Appl. Inf. Syst.*, vol. 10, no. 4, pp. 29–34, Jan. 2016, doi: 10.5120/ijais2016451492.
[17] V. Van Vlasselaer *et al.*, "*APATE*: A novel approach for automated credit card transaction fraud detection using network-based extensions," *Decis. Support Syst.*, vol. 75, pp. 38–48, Jul. 2015, doi: 10.1016/j.dss.2015.04.013.
[18] A. Gupta, D. Kumar, and A. Barve, "Hidden Markov Model based Credit Card Fraud Detection System with Time Stamp and IP Address," *Int. J. Comput. Appl.*, vol. 166, no. 5, pp. 33–37, May 2017, doi: 10.5120/ijca2017914060.
[19] K. Mule and M. Kulkarni, "Credit Card Fraud Detection Using Hidden Markov Model (HMM)," . *India.*, vol. 1, no. 6, 2014.
[20] K. Nian, H. Zhang, A. Tayal, T. Coleman, and Y. Li, "Auto insurance fraud detection using unsupervised spectral ranking for anomaly," *J. Finance Data Sci.*, vol. 2, no. 1, pp. 58–75, Mar. 2016, doi: 10.1016/j.jfds.2016.03.001.
[21] "Kaggle: Your Machine Learning and Data Science Community." Accessed: Apr. 03, 2025. Available: https://www.kaggle.com/