Journal of Harbin Engineering University Vol 44 No. 9
ISSN: 1006-7043 September 2023

Identification Of Vulnerability During Cross Border Transaction in IoT

R. Lingeswari[1], Dr. S. Brindha[2]
[1]Ph.D Research Scholar, Computer Applications
[2] Associate Professor, Department of Computer Science and Applications
[1&2] St.Peter's Institute of Higher Education & Research, Chennai, Tamilnadu, India
lingeswarir.21rsca002@spiher.ac.in ,brindhas.mca@spiher.ac.in

Abstract
The revolution of Internet of Things (IoT) has been triggering demands for the IoT devices in market. IoTs are actively utilized in various social activities that enables the concept of industry 4.0 ecosystem. Conventional models for the detection of traditional models involves the vulnerability analysis, where the decision is carried out using rules embedding into the models. The fraudulent behavior is not reported in case of frequent transactions across cross-borders. In this paper, we develop a machine learning model that is framed as a predictive big data analytics model that solves the problems associated with vulnerability in transactions across cross-border. The study takes into concern various business problems by banks associated with cross-border transactions with its historical data. The machine learning models help banks to captures the details of fraudulent behavior in transactions. The simulation for predictive learning is induced by the machine learning algorithm that uses historical data logs to train the classifier and thereby a model is developed to predict the fraudulent transactions. The simulation results show that the proposed method enables better assessment on finding the vulnerability than the existing methods.

Keywords: Internet of Things, Social Activities, machine learning, predictive learning.

1. Introduction

A recent study conducted in the United States found that those who use mobile payment systems are more likely to mismanage their funds than those who do not [1]. Worldwide, there is a growing trend away from utilizing physical currency in favor of digital financial products [2]-[3]. It is possible that without this knowledge, it will be impossible to devise strategies to eliminate this financial malfeasance [4]-[6].

There is reason to expect that customers will be able to exert more control over their own financial life as a direct result of technological developments in the supply of financial services around the world [7] [8]. Contrary to popular belief, research conducted in the United States and cited above suggests that using some forms of digital payment may potentially raise financial risk.

Due to the anonymity and convenience of these payment options, some customers may be more prone to give in to impulsive purchases than they would be when using cash or credit cards. Therefore, more investigation into how consumers employ these technologies and how they affect consumer financial habits is necessary [9]. The potential for harm to consumers or segments of consumers is one area that needs to be examined so that service providers and policymakers may plan for unfavorable consequences [10].

When it comes to financial user terminals, attempts to identify fraud are directed at the client system. An Internet of Things (IoT) transaction is considered fraudulent if the user hardware identifiers, network usage data, and software activity data are used in a way that is inconsistent with the typical transaction user terminal [7]. Forensic malware analysis tools will be used to collect data from user terminals for study. Methods from malware forensics can be used to mimic the user behavior in order to steal live data from their system [16]. However, just the most crucial information should be gathered, and it should be streamlined so that rapid analysis can be performed, as financial transactions must be processed in real time. Setting the settings outside of the normative past behaviors for the transaction is crucial. The importance of detecting prior information change for transactions between user activities cannot be overstated [13].

In this paper, we develop a machine learning model that is framed as a predictive big data analytics model that solves the problems associated with vulnerability in transactions across cross-border. The deep belief

network classifiers are used to classify the fraudulent behavior in transactions conducted only from the input IoT devices.

2. Related works

Most studies look on how border controls affect economic growth [14]-[17]. In the past several years, the number of cross-border transactions has increased due to the widespread availability of increasingly sophisticated forms of financial technology (FinTech). The increase in output can be traced back to advancements in financial technology, such as streamlined payment methods and reduced fees for online transactions [18]-[20]. Countries of origin and countries of destination are able to conduct financial transactions thanks to remittances from migrant workers, which is becoming an increasingly important role in the economies of many countries. In this article, we look into the challenges faced by authorities as they try to keep an eye on the remittances industry for signs of illicit activity.

Financial risk management in multinational corporations is the focus on cross-border transaction like cross-border e-commerce logistics supervision system (CBELSS), Cross Border e-commerce system (CBES) and cross border transaction with big data (CBTBD). In recent years, the degree to which the Chinese and global economies are interwoven has increased as a result of fast economic progress on both sides. Many Chinese domestic companies are actively seeking opportunities to become international organizations and expand their operations outside of China as a result of the rapid expansion of China social economy. However, MNCs need to improve their understanding of financial hazards and their ability to minimize such risks in order to produce, operate, and expand in a manner that is less risky [16].

The current status of the group company worldwide activities is discussed, and then the underlying challenges that are holding them back are separated. This study screens and ranks the financial risk, determines the likelihood index, severity index, and financial risk index associated with this risk, and provides policy suggestions based on these findings [20]. The purpose is to find the dangers waiting in the international exchange of financial derivatives; it finds and fixes the mistakes that are made during this process; it calculates the possibility index, the severity index, and the financial risk index related to this danger; it screens and ranks the dangers; and it suggests a method for determining which dangers are the gravest. In order to achieve the characteristic identification of transaction risks, it is necessary to first calculate the gap between the ideal solution and risk element. Such behaviour enables the study to accomplish goal of successfully identifying transaction risks[16].

3. Proposed Method

An outline of the methodology is depicted in Figure 1.The study describes the model for vulnerable transactions in cross border transactions carried out via IoT devices.

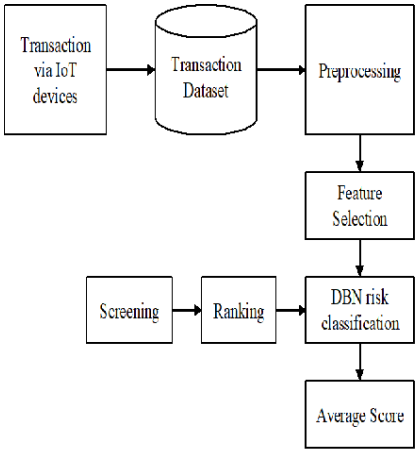

Figure 1: Proposed Model

Data Collection:

The Internet of Things (IoT) is a growing system of interconnected computing devices that can instantly exchange and process data using onboard sensors. The Fourth Industrial Revolution is another name for the IoT. The financial technology industry has received a significant boost from the Internet of Things in recent years, especially in the fields of data security and monetary transaction processing. A mobile POS system is a portable point-of-sale terminal. It is a portable POS system that uses a device that incorporates sensors and other Internet of Things technical processes, allowing it to be used in any location. By simplifying data collecting and exchange, the Internet of Things is assisting the banking industry in making more educated decisions on investments, insurance premiums, customer risks, and other areas of concern.

Pre-processing:

Incorrect data rows are removed as the first step in the pre-processing phase. This action is taken in relation to the data set. The correct format for each variable should be determined after checking with experts in the respective domains. Incorrect samples are frequently deleted because of software, programming, or data entry

errors. It is unfortunate that these slip-ups are frequently exaggerated to avoid further scrutiny. So, they need to be flagged as suspicious right away.

Feature Extraction

The study collect all user-level financial activity and generate a vector set representing it. Looking back at the fraudulent schemes outlined in the previous section, we find that they all exhibit common characteristics, such as an excessive number of transactions, unusually particular destinations, and consecutively big amounts with very short periods in between. These are only some of the features that emerge from our examination of these plans. Features such as recency, frequency, and monetary variables (RFM) are widely used in the business world. If any of these metrics shows abnormal values, an alert needs to be sounded. Combining these numbers yields a wide variety of sequences of varying lengths. Loss of information is inevitable when these sequences are combined into a single dataset via compression. When representing a time series, it is preferable to utilize more than one statistical measure, such as the median, the mean, or the standard deviation.

When working in an unsupervised setting, most feature selection strategies won't work. Simply counting the number of values that are contained within each feature across all of the samples is the simplest test that can be performed on the received dataset. Empty columns should be reorganized or removed altogether because they do not contribute to any analysis or conclusion. The similar approach should be taken with features that always have the same value, as these are easy to spot anytime there is just a little amount of variance in the column.

In addition, we utilize the Pearson correlation coefficient to assess the degree to which a given pair of qualities are related to one another. To determine whether or not a column always returns a value between -1 and 1, we can calculate this metric for every possible column-pair combination and then remove the columns that always return negative or positive values.

A subset of the columns in the dataset may be missing values. In certain cases, a sample may be invalid if its validity depends on the presence or absence of a value for a particular characteristic; in others, the absence of a value for a particular characteristic may not render the sample invalid. Anomaly detection techniques can be weakened by missing data. For missing data, we suggest using the median of the missing column(s). The performance of anomaly detection systems may suffer if the intervals used to define the characteristics are highly variable. First, we divide the value in each column by the standard deviation, and then we subtract the mean from the data. After that point, subsequent feature values will also be condensed to fall inside [1, 1].

Using the scikit-learn machine learning framework available implementations, we train the aforementioned anomaly detection algorithms with the supplied dataset. Some of the model parameters must be set to account for the specifics of each situation.

Bayesian Information Criterion (BIC) plots are generated for relevant models by using the method with different values for this parameter. The BIC relevance cannot be grasped in isolation, but it may be used to compare and contrast the strengths of different models. Find the inflection point when the curve stops reducing noticeably to create the threshold for the number of components. One method used commonly in cluster analysis is known as the elbow technique.

3.1. DBN Classification

Unsupervised training of a deep belief neural network (DBN) could lead to the network eventual ability to probabilistically recreate the results of its inputs. In later steps, the layers are converted into feature detectors. After this phase of training is complete, a DBN can be taught to do classification tasks under human guidance. DBNs are constructed using a large number of unsupervised, fundamental networks like restricted Boltzmann machines (RBMs).

When one hidden layer is passed on to the next, it becomes the sub network visible layer. Using generating energy and connecting layers but not their individual nodes, RBMs are a subset of undirected models. It takes input from two layers, one of which is transparent and the other hidden. Applying contrastive divergence to each subnetwork provides an unsupervised layer training from the lowest layer pairs as in Figure 2.

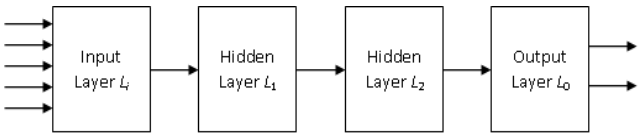

Figure 2: DBN architecture with a output layer and 2 hidden layers

To increase the process of the learning curve associated with RBM, the contrastive divergence approach is

employed. This approach is predicated on the idea that all hidden units can be updated simultaneously while the visible ones are first updated, reconstructed from the hidden ones, and then updated again. Binary units with arbitrary weights are joined at various levels to form a simple belief network. The belief network is an acyclic graph, which means that the information types it trusts may be seen at its leaf nodes. In order to produce outputs that are consistent with observations, a DBN must observe the unobserved stochastic units and hence based on which the weights should be modified. A node in a DBN can have a value of 0 or 1, and the likelihood of it switching to 1 is determined by the node bias and the quantity of weighted information it receives from other nodes as expressed below:

$$w_{ij}\left(t+1\right)=w_{ij}+\eta\frac{\partial\log\left(p\left(v\right)\right)}{\partial w_{ij}}$$

Here

η - learning rate

$p(v)$ - visible vector probability

j – visible layer,

i - hidden layer, and

W - weight matrix between j and i.

To begin training a DBN, a layer of features based on observable units must be learned using the contrastive divergence (CD) technique. The next phase involves feature learning from the final hidden layer while considering the outputs of activation functions from previous hidden layer and visible units. After the last hidden layer has been trained, the entire DBN is evaluated. The greedy learning technique can be used as a starting point for training a DBN. This is because each layer of the training RBM is optimized with the CD approach, and the succeeding stacking RBM layer uses these values to optimize itself. It is feasible that the global optimum is necessary since each layer is learnt iteratively to obtain optimal values.

We choose samples created by persistent contrastive divergence because it produces samples that are heavily weighted toward more recent data. We can then employ samples that are heavily weighted toward more current information. Starting from a random seed at the beginning of each iteration, samples are generated for the model distribution that are sampled on the hidden states. These latent states are preserved from one data point to the next.

By treating the hidden layer, DBNs can be constructed from a stack of RBMs. Building a DBN is the end consequence of this procedure. Applying the previously mentioned, unsupervised RBMs as well as the addition of hidden layers has been shown to raise the lower bound on the log-likelihood of the training data. The possibility of encoding abstract features at higher levels is higher since these attributes tend to appear there, and they are thus very useful for classification applications. Back propagation of errors is used to fine-tune the entire DBN network once the top layer has been trained using supervised learning techniques. DBNs facilitate the merging of disparate data sources into unified wholes. It is feasible to produce a new association layer by constructing a shared visible layer at the top of both preprocessing hierarchies, which are constructed independently for the two inputs. This paves the way for constructing preprocessing hierarchies for both inputs. Therefore, DBNs may store not only flat data, but also data organized in hierarchical structures like trees.

It is not only costly and time-consuming to manually label data, but it is also a tedious and tedious endeavor. Data labelling could be increased up with the use of pseudo labelling, as demonstrated in Figure 3.

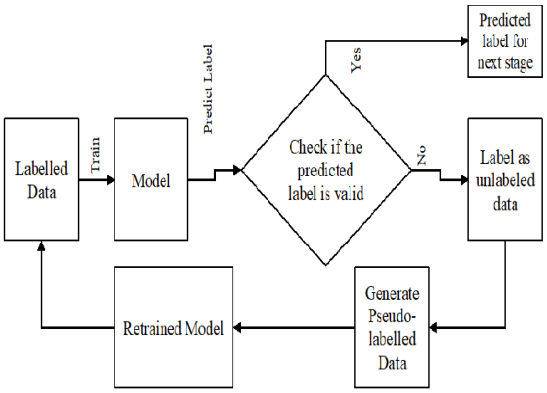

Figure 3: Labelling method

Screening the CBT

Evaluating the implications and potential of the identified hazards, and assessing the prospective consequences on the project objectives in descending order of significance, are all part of the CBT derivatives design (Figure 4). The final product is a document that includes both a list of potential threats and a plan for dealing with them. Ultimately, this report will be used to prioritize risks based on their possible impact on the project desired outcomes. The project overall risk can be assessed by comparing the weights given to the various risks. The first stage is to identify the risks connected with CBT derivatives transactions and the final stage is to screen

and classify the transaction risks associated with such trades.

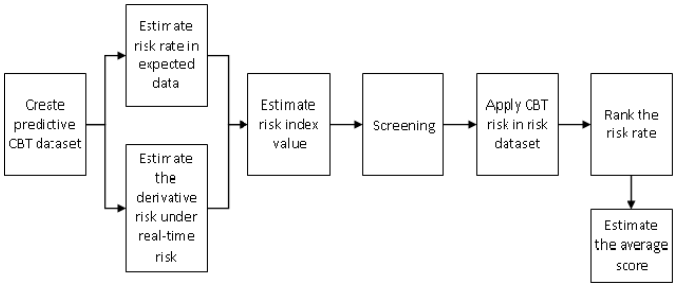

Figure 4: Process of Risk estimation

Transaction risk connected with international financial derivatives can be screened and ranked using the likelihood index, the severity index, and the financial risk index. Utilizing these indicators is important. Risk connected with foreign transactions involving financial derivatives is often described using the following formula for the possibility index of transaction risk:

$$\delta_l(s) = \max\left\{P_r\left(s > s_{\max}\right), P_r\left(s < s_{\max}\right)\right\}$$

where

Pr(·) - probability value, and

S- trading node of CBT derivatives.

Cross-border transaction in financial derivatives can be characterized by the following formula, which accounts for both the frequency and magnitude of trading hubs.

$$\delta_e(S) = \max\left\{\frac{|s|-|s_{\max}|}{|s_{\max}|}, \frac{|s|-|s_{\min}|}{|s_{\min}|}\right\}$$

To obtain the solution for identification of risks associated with the CBT derivatives in cross-border operations, we employ the analytical hierarchy process to screen the risks of financial derivatives in cross-border operations and to delete risks with minimal impacts on CBT derivatives. This is done to cut down on the amount of potential transaction hazards that need to be found. The dangers of engaging in international trading in financial derivatives can be broken down into the following categories:

$$\alpha = P/P_{max}$$

where

P_{max}- financial risk limit of derivatives associated with CBT,

P- financial risk in real-time for the derivatives associated with CBT, and

α- rate of financial risk.

Financial derivatives transaction risk DRI is expressed as a percentage, and it is used as a screening and ranking index for cross-border business deals. Multiplying the anticipated risk in CBT derivatives using actual risk provides the value of this index.

$$DRI=P \times \alpha$$

Figure 2 depicts the method used to screen and evaluate dangers related to overseas deals using financial derivatives. Analytic hierarchy technique is used to calculate the severity, probability and financial risk index of CBT derivatives. The scale of the problem that needs to be addressed to jointly identify transaction risk in cross-border financial derivatives is then reduced by screening the risks involved.

Finding Risk Derivatives

The analytic hierarchy approach is used to address the distance between indicators of the CBT derivatives risk in cross-border transactions, which is a statistical problem. This method is used to limit the mutual interference (MI) between various indicators and to avoid the dimensionality problem between the various indicator properties.

Assuming the identifiers are a multi-attribute vector x= (x_1, x_2,...,x_p) with a covariance matrix $s$anda mean value μ= (μ_1,μ_2,...,μ_p),, we can compute the Mahalanobis distance of the identifier index as follows.

$$M(x) = \sqrt{(x-\mu)^T s^{-1}(x-\mu)}$$

Let

A_i- index value of i^{th} transaction risk element,

$B^+$and B^-- positive and negative ideal solution,

s^{-1}- inverse covariance matrix for a transactional risk.

Since s^{-1} is linearly invariant, it is unaffected by changes in the identification indicator dimensions, as well as the correlation between the indicator qualities has been broken. Moreover, the connections between the various characteristics of identifiers have been severed. The transaction risk component is defined as the difference between the B^+ to B^-.

$$M\left(A_i, B^+\right) = \sqrt{\left(A_i - B^+\right)^T s^{-1}\left(A_i - B^+\right)}$$

$$M\left(A_i, B^-\right) = \sqrt{\left(A_i - B^-\right)^T s^{-1}\left(A_i - B^-\right)}$$

Considering the proximity element of the CBT risk is c_i, the following formulation is attained:

$$c_i = \frac{M\left(A_i, B^-\right)}{M\left(A_i, B^-\right) + M\left(A_i, B^+\right)}$$

For financial risks, the positive and negative distances from the risk element to the ideal solutions are defined by the Mahalanobis distance. To model the risk characteristic

of CBT, a combined risk identification model is built. This is achieved by estimating the proximity of the risk component of international trading in financial derivatives.

Classification Score

The system calculates a score for each individual. It is difficult to make direct comparisons between these numbers because they are measured on such different scales. The formula we use to give the highest possible score of 1 to the most suspicious person in the dataset and the lowest possible value of 0 to the most ordinary person is as follows:

$$\text{scorenew} = [score_{old} - \min(score_{old})] / [\max(score_{old}) - \min(score_{old})]$$

The next step is to calculate an average of the first three scores in order to produce a fourth score for each data point. It is possible that our methods are the best fit for the data via feature extraction, if the histograms obtained are different from one another. The average score and the other outcomes should be fairly consistent with one another, although the other possibilities may differ. In order to set a higher accuracy, it is sometimes necessary to average scores from multiple sources. The most suspicious accounts are given top priority for further human investigation with the help of money laundering and financial professionals.

4. Results and Discussions

Experiments are conducted to test the viability and efficacy of a method for identifying risk features in cross-border financial derivatives deals. The first stage in getting sample data for use in a transaction risk analysis is classifying the raw data, choosing samples at varying risk levels, and collecting the data that is produced as a result. The sample data must be separated into a training data set and a test data set before the experiment can be run. It is recommended that there be 5,000 training samples in the training data set and 1,000 test samples in the test data set.

Since there exist no cross-domain CDT datasets in real-world, we modelled a transactions system, wherein we transferred payments via IoT across different servers. The log results are collected and stored in the form of a database and then we processed the financial risk across CDT. The simulation is conducted in terms of how well the system identifies the potential risk and discards the transactions.

The simulation is modelled in python and the results show that the proposed method achieves higher rate of accuracy in detecting the risk in training and in testing modes. Likewise, it is seen that the proposed method has higher precision, recall and f-measure rate in training and testing modes.

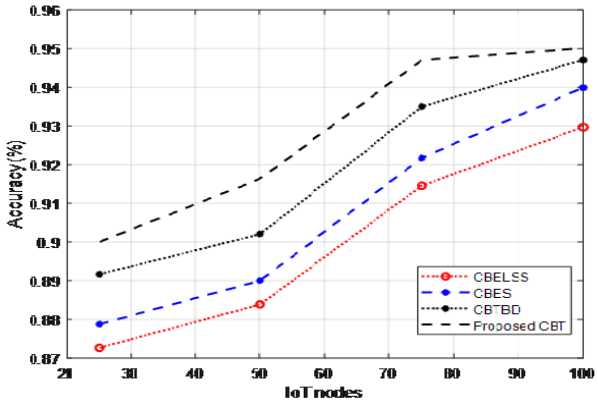

Figure 5: Accuracy

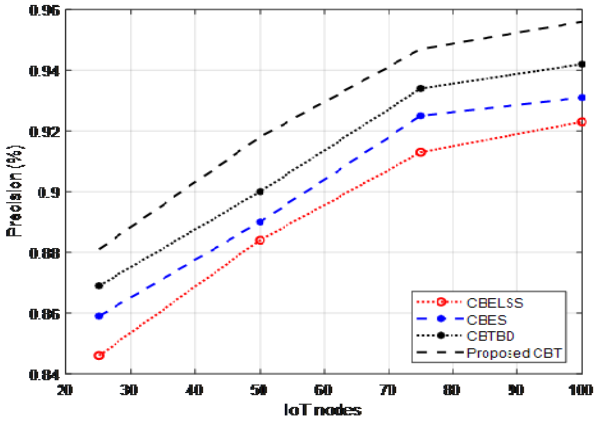

Figure 6: Precision

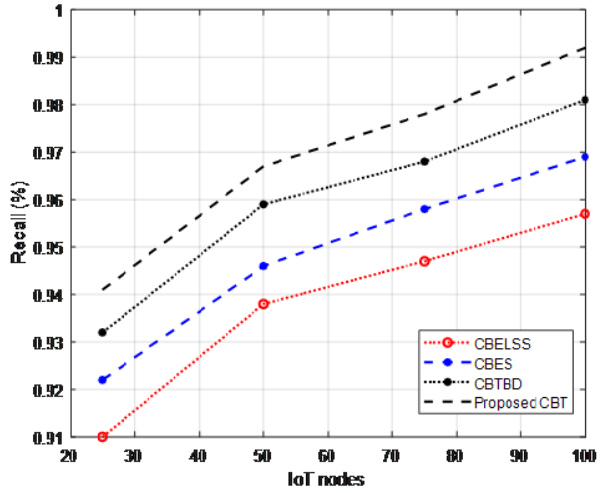

Figure 7: Recall

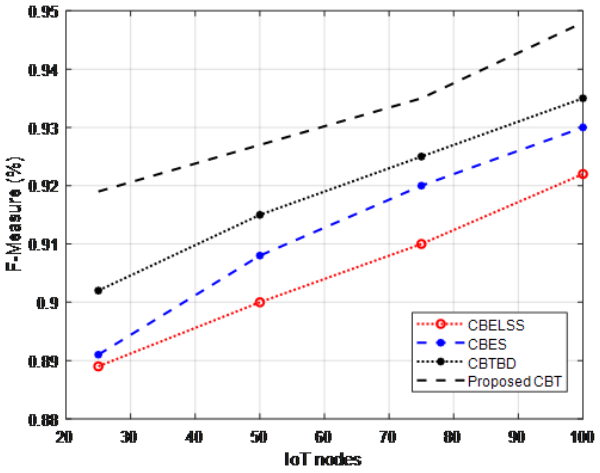

Figure 8: F-Measure

5. Conclusions

We address the challenges of transacting across international borders by creating a machine learning model in the form of a predictive big data analytics platform. This research examines a wide range of banking issues concerning cross-border transactions by analyzing historical data. Machine learning algorithms aid financial institutions in the detection of fraudulent transactions. The manageability of the financial derivatives trading firm can be enhanced by employing the simulation management data identification and query approach to more precisely identify the risks associated with CDT derivatives. It is necessary to have a more in-depth understanding of the potential threats. A model that can foresee fraudulent transactions can be generated using a simulation generated using a machine learning technique that employs the utilization of previous data logs to train a classifier. Then, you can apply this model. From what we can tell in the simulation, the proposed method does a better job of assessing susceptibility than the methods that are currently being employed.

References

[1] Binder, A., Jadhav, O., & Mehrmann, V. (2021). Model order reduction for the simulation of parametric interest rate models in financial risk analysis. *Journal of Mathematics in Industry*, *11*(1), 1-34.

[2] Zhu, L., Li, M., & Metawa, N. (2021). Financial risk evaluation Z-score model for intelligent IoT-based enterprises. *Information Processing & Management*, *58*(6), 102692.

[3] Fang, L., & Jianyuan, G. (2021, March). Internet Financial Risk Analysis and Supervision Suggestions. In *2021 2nd International Conference on E-Commerce and Internet Technology (ECIT)* (pp. 446-449). IEEE.

[4] Qu, M., & Li, Y. (2021). Financial risk early-warning model based on kernel principal component analysis in public hospitals. *Mathematical Problems in Engineering*, *2021*.

[5] Moslehpour, M., Al-Fadly, A., Ehsanullah, S., Chong, K. W., Xuyen, N. T. M., & Tan, L. P. (2022). Assessing financial risk spillover and panic impact of COVID-19 on European and Vietnam stock market. *Environmental Science and Pollution Research*, *29*(19), 28226-28240.

[6] Cao, Y., Shao, Y., & Zhang, H. (2022). Study on early warning of E-commerce enterprise financial risk based on deep learning algorithm. *Electronic Commerce Research*, *22*(1), 21-36.

[7] S. Brindha and N. K. Sakthivel, "Virtual Machine Dynamic Migration Strategy based Intelligent Flow Forecast Technique for Cloud Data Centers", International Journal of Emerging Technologies and Innovative Research (IJETIR), Volume 6, Issue 4, Pp 368-376, April-2019.

[8] Hubert, R., Marginean, I., Cardona, M., Clapp, C., & Sillmann, J. (2021). Addressing challenges of physical climate risk analysis in financial institutions. *institutions*.

[9] Cui, Z., An, F., & Zhang, W. (2021). Internet financial risk assessment based on web embedded system and data mining algorithm. *Microprocessors and Microsystems*, *82*, 103898.

[10] Choi, T. M. (2021). Risk analysis in logistics systems: A research agenda during and after the COVID-19 pandemic. *Transportation Research Part E: Logistics and Transportation Review*, *145*, 102190.

[11] Zhu, L., Li, M., & Metawa, N. (2021). Financial risk evaluation Z-score model for intelligent IoT-based enterprises. *Information Processing & Management*, *58*(6), 102692.

[12] Wang, C., & Liu, S. (2021). Innovative risk early warning model based on internet of things under big data technology. *IEEE access*, *9*, 100606-100614.

[13] Shang, H., Lu, D., & Zhou, Q. (2021). Early warning of enterprise finance risk of big data mining in internet of things based on fuzzy association

rules. *Neural Computing and Applications*, *33*(9), 3901-3909.

[14] Hongjin, S. (2021). Analysis of risk factors in financial supply chain based on machine learning and IoT technology. *Journal of Intelligent & Fuzzy Systems*, *40*(4), 6421-6431.

[15] Zeng, G., Chi, J., Ma, R., Feng, J., Ao, X., & Yang, H. (2022). ADAPT: Adversarial Domain Adaptation with Purifier Training for Cross-Domain Credit Risk Forecasting. In *International Conference on Database Systems for Advanced Applications* (pp. 353-369). Springer, Cham.

[16] R.Lingeswari and Dr.S.Brindha "A Review on the secured transitions in Financial Instituition using IoT Big Data", International Journal of Early Childhood Special Education (INT-JECSE) ISSN: 1308-5581 Vol 14, Issue 03 2022

[17] Zhang, Y., Wang, Y., Ahmad, A. B., Shah, A. A., & Qing, W. (2021). How Do Individual-Level Characteristics Influence Cross-Domain Risk Perceptions Among Chinese Urban Residents?. *SageOpen*, 215824402110035

[18] Chen, J., Zhan, Z., He, K., Du, R., Wang, D., & Liu, F. (2021). XAuth: Efficient privacy-preserving cross-domain authentication. *IEEE Transactions on Dependable and Secure Computing*.

[19] Rowson, T. S., Meyer, A., & Houldsworth, E. (2022). Work identity pause and reactivation: a study of cross-domain identity transitions of trailing wives in Dubai. *Work, Employment and Society*, *36*(2), 235-252.

[20] R.Lingeswari and Dr.S.Brindha " Machine Learning-Money Transaction", National Conference on Learning Techniques in Artificial Intelligence Computer Science,Computer Application and Mathematics. ISBN No. 978-93-91535-08-7, April 2022.

[21] Shou, Y., & Olney, J. (2022). Measuring risk tolerance across domains: Scale development and validation. *Journal of personality assessment*, *104*(4), 484-495.

[22] Zhang, J., & Tan, R. (2022). Radical Concept Generation Inspired by Cross-Domain Knowledge. *Applied Sciences*, *12*(10), 4929.

[23] Prasat, K., Sanjay, S., Ananya, V., Kannadasan, R., Rajkumar, S., Raut, R., & Selvanambi, R. (2022). Analysis of Cross-Domain Security and Privacy Aspects of Cyber-Physical Systems. *International Journal of Wireless Information Networks*, 1-26.

[24] Sun, P., & Gu, L. (2021). Optimization of cross-border e-commerce logistics supervision system based on internet of things technology. *Complexity*, *2021*.

[25] Protopappas, L., Sideridis, A. B., & Yialouris, C. P. (2020). Implementation Issues of Cross Border e-Government Systems and Services. In *HAICTA* (pp. 155-166).

[26] R.Lingeswari and Dr.S.Brindha " Analysing and classification of Attacks in Financial Transactions using Machine Learning", Inter National Conference on Innovative Technologies and their Applications in Higher Education-Science. ISBN No. 978-93-92042-31-7, October 2022.