ISSN 2278-3091

Volume 9, No.5, September - October 2020

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse248952020.pdf

https://doi.org/10.30534/ijatcse/2020/248952020

ANOMALY DETECTION IN DOCUMENT VERIFICATION SYSTEM USING DEEPLEARNING IN HYPERLEDGER

Priya N[1], Dr. Ponnavaikko M[2] , Rex Aantonny[3]

[1]Research scholar, Department of computer science, Bharath Institute of Higher Education And Research, Chennai, India, priyabiher@gmail.com

[2] Provost, Bharath Institute of Higher Education And Research, Chennai, India, ponnav@gmail.com

[3] Founder & CEO, Rex Cyber Solutions Pvt Ltd, Bangalore, India, rex@rexcybersolutions.com

ABSTRACT

The educational industry is being integrated with technology and it has raised various challenges in maintaining the documents of academic details for each candidate in a longer period. The individual seeks higher education or recruiting for any industry then the verification of documents holding academic details is essential as well as to provide a reliable solution to avoid any academic fraud. Thus verification of certificates is done by blockchain in a completely decentralized transparent manner. Hyperledger Indy provides a digital identity that is rooted in blockchain for all records involved in the verification process. Records used for verification must not hold any anomalies to provide accurate data. Detecting anomalies is essential and increases the efficiency of the system. In this paper, an anomaly detection model is proposed for academic records data using deep learning by Keras on the top of the TensorFlow. Experimentation involves deep learning algorithms using the facial recognition model specifying some features like image size, facial weight, and width processing time all that are considered. The test results show that the proposed system model provides high performance, with high efficiency and low cost includes the minimum amount of processing time. By detecting, the anomalies using ML algorithms assure the trustworthiness of the documents involved with more transparent transactions.

Key words : Anamoly detection, blockchain, hyperledger Indy , deep learning techniques, Tensorflow

1. INTRODUCTION

The educational industry growing rapidly with the development of technology. All educational institutions have the practice of maintaining the data of each candidate digitally now. When the candidate applied for a new job or higher education the verification of all academic details are mandatory to avoid any fraudulent or forged data. This verification is done manually by olden days with long time processing like a month also facing issues like loss in records and the fraudulent activities for making forgery certificates. This record-keeping system of the institutions is inefficient and wastes money and time for the students as well as employers. Many educational verification services and employee background educational services existing currently but they still have some limitations such as privacy, transparency, high cost, and some scams still exist with third parties[1].Institutions are the centralized authority to verify the records. It will be difficult for the students and employers when they give requests for official verification. The cost for verification involved is high and that will be more burden to users. The privacy and security will be taken care of maintaining the records to avoid risks like data theft and loss, modification. The storage size for handling all records must possess high capacity that also very difficult to manage.

A novel blockchain mechanism-based technique is used for maintaining records and share the records with others. Blockchain technology provides an immutable distributed ledger [2] and the verification done by some consensus algorithms. The decentralized characteristic nature of Blockchain technology is used to maintain the records in a most secure method. Educational institutions and organizations are no need to maintain the centralized storage of all records and reduce the cost for storage and accessing them.

Here Permissioned blockchain networks such as hyperledger are selected for validation mainly because of its cost and performance and the confidentiality of data used. The secure and private nature of blockchain assured the confidentiality of the data stored on it. So the academic data which is stored in hyperledger used for verification must be anomaly free. In case of the irrelevant, duplicate, and forged data could be used for verification then the final results will be affected. To prevent this we are using some deep learning algorithms for detecting anomalies. Anamoly detection depends on the unsupervised methods and finding unknowns from the knowns. An autoencoder neural network in deep learning is mainly used for anomaly detection[3]. The autoencoder algorithm applies the backpropagation method and setting the output values that would be the same as the input values are given.

For building the models used for anomaly detection, a Tensorflow open-source framework for machine learning has been used here. The dataset is used for creating the model using deep learning algorithms for university records which will provide all necessary details of each candidate for background checking.

The motivation of the work is summarized as follows:

- To detect anomalies in an academic records which used as input for verification.
- Deep learning algorithms are used to find the erroneous data by creating the training data set which are helping to find the key features of every file in the dataset. Based on that the behavior of fraudulent data would be analyzed
- Tensor flow open-source tool is used to create the model by accessing, organizing, and processing the data.
- Finally, the dataset used for verification would be clear data without any forge or fraudulent and the accuracy precision and recall could be measured.

This paper is framed as follows. Section II involves related works in anomaly detection by deep learning algorithms in academic records verification in blockchain technology. Section III explains the technical background of blockchain and deep learning algorithms. Section IV provides the methodology used to detect anomalies. Section V gives the experimental setup which supports to implement the methods. and results analysis. Finally, in section VI Conclusion is discussed and References included..

II. RELATED WORKS

This section reviewing about anomaly detection with some machine learning algorithms in the hyperledger.Anomaly detection in the blockchain is mainly used to identify any fraudulent transactions exists. This will be carried out by observing the characteristics and features of transactions involved in the identity management system in the blockchain. Before executing the procedure for validating the academic records, the database must be checked and detecting any anomalies occur. This anomaly detection is done by a few deep learning algorithms. Sreerag Iyer et al[4] proposed a wastewater recycle control system which based on blockchain technology uses some machine learning algorithms like polynomial regression, DBSCAN, autoencoders, and LSTMnetworks to detect the tampering of the data. But storing enormous data in the current blockchain platform and adapting to changing costs make the system inefficient. Blaž Podgorelec[5] framed a machine learning-based method which gives the automated signing for blockchain transactions that involve a unique identification of anomalous transactions that is performed on data from Ethereum public main network. Tsuyoshi Id´ e proposed[6] a framework for collaborative anomaly detection on blockchain for noisy sensor data. Three main problems like validation, consensus creation, and data security are analyzed using machine learning algorithms. Hakim Ghazzai Et al developed [7]a fraud detection system for insurance data based on hyperledger and some machine learning algorithms. XGBoost and VFDT algorithms are used to classify and detect fraudulent data for insurance claims. Davy Preuveneers et al [8] framed a hyperledger based federated learning model which has incremental updates for the anomaly detection which is chained together with the distributed ledger. But their model aggregation algorithms could not satisfy the unreliable network connectivity between connected nodes. M Signorini et al proposed [9] an Anomaly Detection solution for blockchain which used to provide the prediction tools for heterogeneous malicious transactions.

Sirine Sayadi et al created a new model for anomaly detection for bitcoin transactions[10]. The One-Class Support Vector Machines (OCSVM) algorithm to detect outliers and the KMeans algorithm for grouping outliers are used in their model. But minimization of false-positive and detecting other vulnerabilities in blockchain are still under research.The above researches contributed their work towards the fraudulent detection in blockchain transactions based on some deep learning algorithms. In case more nodes participated and the high maintenance cost spent is becoming inefficient in achieving the output. Scalability is the main issue of this research work. Our contribution in this research is to optimize the detecting techniques and provide security and integrity in the verification of documents used.

III.BACKGROUND

To understand the technologies behind our research work it is necessary to know about the basic things about blockchain ,hyperledger Indy and some deep learning methods for anomaly detection in detail.

A.Blockchain

Blockchain technology was framed by Satoshi Nakamoto based on cryptocurrency and bitcoin in 2008.[2] It is an open and immutable distributed ledger that supports the exchange of digital assets without any intermediaries. It keeps all the transactions between two untrustable parties in a verifiable and immutable manner. The main functionalities of Blockchain are reduced cost, increased speed, increased security, reduced fraud, and reduced risk. The rules that can handle the transactions are accepted by the participants of the network by using smart contracts. For validating each transaction , participants of the blockchain network use a consensus mechanism by creating blocks in the chain. Block stores all transactions involved. Every block contains a block header , timestamp, and a summary of total transactions with Merkle root. Every block inherits from the previous block by its cryptographic hash value of the existing block. That will be used to generate the new block’s hash to make the blockchain tamper-proof. There are two types of models in blockchain i.e permissionless and permission. Permissionless framework is applied for cryptocurrency.The permission blockchain is mainly for smart contracts.Consensus is formed based on the maximum supporters with maximum benefits.Each member of the network participated in the consensus and cooperated with the other nodes.

B. Hyperledger Indy

Hyperledger is a modular blockchain framework that was first released in December 2015 by the Linux Foundation.[15] Participants involved in the transactions are known to each other but never have trust upon themselves. Hyperledger supports the smart contracts using some programming languages like Java, Go, and Node.Js. Hyperledger reaches consensus by ordering and validating the transactions. Hyperledger includes distributed ledger frameworks like Sawtooth, Indy, and Ursa. Hyperledger Indy is based on blockchain distributed ledger technology mainly used for decentralized identity. Mainly Indy induced the users to store and maintain the privacy of the data. Decentralized identifiers[DID] are used as the keys here.

C. Anomaly detection In machine learning:

Anomaly detection is an identification mechanism used to find the fraudulent or suspicious items in the data set. It detects the unusual changes in the behavior of the system also predicts the failures and provide the accuracy in the results. Anomaly detection analysis the problems then provide the services to solve that.To find anomalies few deep learning algorithms like Isolation Forest,one_class SVM, Elliptic Envelopes, and local outlier factors are used. When developing the anomaly detection with deep learning techniques an autoencoder is helping a lot to get a correct solution.[12]

Anomaly detection algorithms are classified into two types(i) Outlier detection and (ii)Novelty detection. Outlier detection is a form of unsupervised learning. These unsupervised learning algorithms classify the entire dataset into a standard form of data and anomalies. Novelty detection is based on Supervised learning. These algorithms predict the given input data is in standard form or anomaly. An auto-encoder is one of the types of artificial neural network which helps to get the standard data in an unsupervised manner. The working of an auto-encoder involves compressing the input data and reconstruct the input variables.

D.Tensorflow

Tensor flow is an open-source platform that will help to solve the problems based on machine learning. It helps the researchers to build a neural network model with deep learning techniques. Data will be preprocessing before building a model, after that the training data set is used to estimate the model. Tensorflow's name is derived from the computational matrix called Tensor, All values in tensor hold the identical type of data called shape.

Each TensorFlow operations are held through the graph. Every researchers prefer TensorFlow because of it accessible to everyone. Tensorflow library is used to build the CNN or RNN neural networks. Based on graph computation neural network is constructed and visualized.

IV.METHODOLOGY

In this section, we design the deep learning approach to find anomalies for the usage of the hyperledger blockchain system. The data used for anomaly detection is unsupervised data. It doesn't possess any common patterns among them. A variety of images of documents in the form of pdfs, jpeg files, word files are taken as the input for verification. Due to this non uniformity of data taken took more time with high cost and possessing less accuracy in the results.

The proposed anomaly detection system builds an autoencoder that filters the input data and reconstruct the required data. For creating this framework collecting all kinds of data and some deep learning algorithms on it and recognize the types of images. Tracing out the anomaly in that dataset and creating a new set of images.

A.Systemflow Architecture

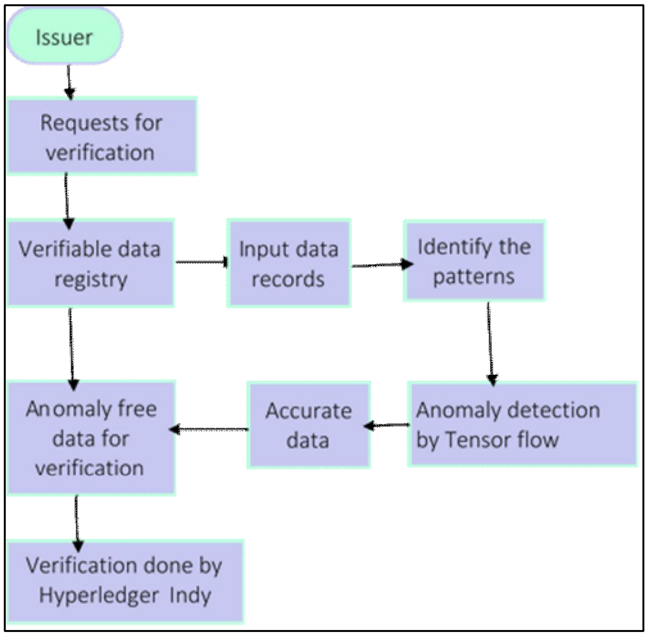
Issuer
Requests for verification
Verifiable data registry
Input data records
Identify the patterns
Anomaly free data for verification
Accurate data
Anomaly detection by Tensor flow
Verification done by Hyperledger Indy

Fig:1 Proposed solution structure

Fig 1 indicates the proposed structure of detecting anomalies in educational records for verification of its originality.

B.Workflow

For document verification, the academic records are given as the input data. But that dataset contains both standard data(inliers) as well as the fake or fraudulent data(outliers). Using some simple deep learning methods could not provide the correct solution in the case of anomaly detection. Here autoencoders are used for this anomaly detection process.autoencoders are the artificial neural networks that would develop a framework and reconstruct the output without anomaly Autoencoders are mainly used to find the patterns of the given input data. It encodes the data and transforms into latent representation.

The academic record data set are divided into testing data and training dataset. There is no anomalous data present in the training data. Next, the output will be error-free while creating patterns. After implementing the autoencoder the data set with labels with standard and anomaly have been discarded to make the set prepare for the unsupervised machine learning. By visualization, the reconstruction of the output image would have been created from the given input image by using Tensorflow By calculating the mean square error of the input images from the predicted output from autoencoder it is easy to identify the anomalies and outliers.

V. EXPERIMENTAL SETUP &RESULTS

Memory	vCPUs	Transfer	SSD Disk
16GB	6vCPUs	6TB	320GB

Table:1 System requirement for implementation

Table 1 describes the hardware, software requirements for implementation of our research work.

S.no	Proposals per minute	Man power for issuing response	Cost approx .(USD)	Response time	Network
1	100	NIL	100	<1sec	LTE
2	200	NIL	100	<1sec	LTE
3	500	NIL	100	<1sec	LTE
4	750	NIL	100	<1sec	LTE

Table 2: Data taken for verification

Table 2 denotes the proposals given for verification and approximate cost of 100USD implemented in LTE network system. Documents taken for verification undergone an anomaly detection are taken as input proposals and it compressed into latent space representation. Using tensor flow the input data are reconstructed from the latent representation. The above tables explained the experimental setup details as well as the proposals given for verification with outlier data.

In this implementation of document verification inside hyperledger Tensor flow is used to find detecting the anomalies. For this implementation, a facial recognition model is created in the hyperledger backend system that is used to recognizing the face using TensorFlow. Here the factors taken for the experiment are processing time, image size, and width of specific facial features. The training dataset is used to create the backpropagation model using machine learning algorithms. By analyzing the factors of facial recognition processing time, and image size the results produced are less a second.

Hashing	Time for 2 million matching	Matching per 1sec	Size
48bit Length	0.0098 sec	46 million / sec	35 MB

Table:3 Performance results

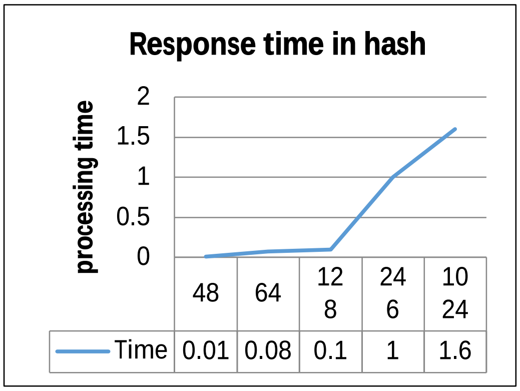

Fig:3 requests vs response time(in hash values)

Table 3 shows the recognition processing time that is lesser than a second even though the number of requests is increasing.If any malicious activities like forging or altering the data occur it will be verified by the proposed model and the processing time of the image recognition would be affected. The processing time values are saved in its hash value. By verifying the images the machine learning algorithms create the backpropagation model and check for any anomaly. Figure 3 shows the change in hash values that indicates the presence of malicious activities in the system. Our results clearly have shown in figure 4 denote the increase in efficiency and throughput with low cost and minimum time. The proposed system provides the solution to fake document issues in the hyperledger document verification system.

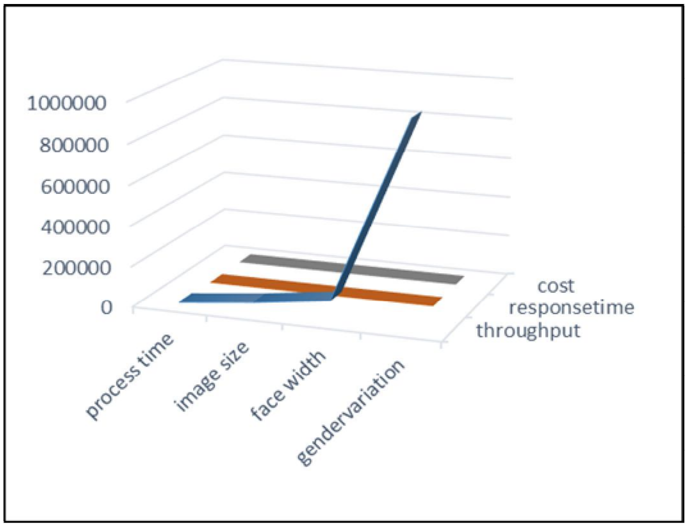

Fig 4: performance results

V.CONCLUSION

This research work presents a solution to the risks that occurred during the document verification in the educational industry. The existing system is a very vast and complex process while performing for an enormous number of requests. This paper has done a groundwork on hyperledger based anomaly detection using TensorFlow yields better results.

Our Experimentation results proved that the efficiency of the system had improved with low cost and minimum processing time. This framework provides high throughput and reduces the risks that happened in the verification system. Using this method in real-time applications and interrogate with the existing verification methods would make a great revolution in the educational industry without any scam and make the system as a trustable one. Future research is applying this model to the supply chain industry in detecting and avoiding the information fraud.

ACKNOWLEDGEMENT

The authors would like to thank Keerthana Bose, Senior Developer at CGI, Virginia, USA for her financial support and Rex Aantonny CEO of Rex cybersecurity solutions Pvt ltd for giving technical support to complete this research..

REFERENCES

1. Raghav et al "**Tamper-Proof Certificate Management System**", IEEE Conference on Information and Communication Technology 2019

2. Satoshi Nakamoto, "**Bitcoin: A Peer-to-Peer Electronic Cash System**",2008.

3.https://dzone.com/articles/dive-deep-into-deep-learning-using-h2o-1

4. Sreerag Iyer, Snehal Thakur, Mihirraj Dixit, Rajneesh Katkam, Ashish Agrawal, Faruk Kazi, "**Blockchain and Anomaly Detection based Monitoring System for Enforcing Wastewater Reuse**" ICCCNT 2019.

5. Blaž Podgorelec , Muhamed Turkanovic´ and Sašo Karakaticˇ, "**A Machine Learning-Based Method for Automated Blockchain Transaction Signing Including Personalized Anomaly Detection,**" https://doi.org/10.3390/s20010147,december 2019.

6. Tsuyoshi Id´ e, "**Collaborative Anomaly Detection on Blockchain from Noisy Sensor Data**" doi: 10.1109/icdmw.2018.00024,2018

7. Hakim Ghazzai ,Najmeddine Dhieb, ,Hichem Besbes, And Yehia Massoud, "**A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement**", april 2020.

8. Davy Preuveneers,Vera Rimmer,Ilias Tsingenopoulos ,Jan Spooren ,Wouter Joosen and Elisabeth Ilie-Zudor, "**Chained Anomaly Detection Models for Federated Learning: An Intrusion Detection Case Study"**, doi:10.3390/app8122663.

9. Priya N,Dr.M.Ponnavaikko,Rex antony, **"blockchain based validating documents by self soverign identity using hyperledger Indy**",

https://www.doi.org/10.37896/GOR33.03/493

10. M Signorini, M Pontecorvi, W Kanoun, "**BAD: a Blockchain Anomaly Detection solution**" https://arxiv.org/abs/1807.03833v2,july 2018.

11. Sirine Sayadi, Sonia Ben Rejeb, Zièd Choukair., "**Anomaly Detection Model Over Blockchain Electronic Transactions**" 15th International Wireless Communications and Mobile Computing Conference,2019.

12. Priya N, Dr.M.Ponnavaikko, Rex antony , "**Analysis of Block Chain based E-Procurement System –A Novel Approach"** Advancement in Engineering, Science & Technology J. Mech. Cont.& Math. Sci,August (2019) pp 145-155.