GIET UNIVERSITY JOURNAL, 2022, VOLUME 1, ISSUE 1.
Manuscript received on 2nd July 2022: Revised on 16th July, 2022; Accepted on 26th July, 2022

Credit Card Fraud Detection using Logistic Regression and Data Visualization using PowerBI

Dr.M.RamKumar, Associate Professor,
HKBK College of Engineering.
Chaithanya K, Snehalatha A,
Supriya R, Veena J, Department of computer science and engineering

***Abstract*— The project speaks about fraud detection. Frauds can be of many kinds say manual fraud, Customer detail stolen, Debt-card frauds, credit-card fraud, online authentication details stolen fraud and many more. This project makes an attempt to solve the fraud that happens during a credit card transaction. During this century credit cards have become a most common means of pay. Banks suggest its customers to take up credit cards for their day to day living as it plays a very crucial role in one's financial crisis.**

These kind of frauds goes on increasing proportional to the amount of people using credit cards. In order to reduce this many attempts were made having the previous fraudulent transaction as a base. The upcoming transaction datasets are being used into trained models such as KNN, Logistic Regression or Random-Forest.

***Keywords*— credit card fraud, Fraud Detection, Data mining, real-time fraud transactions, Logistic Regression, test-train-split, data visualization, powerbi.**

I. INTRODUCTION

IN recent years we know that online payment methods are used widely which in turn increased the non-cash electronic transactions. Credit cards represent one of the most common electronic payment methods. It has a structure in such a way that it looks thin, rectangular or box in shape made of plastic and its details written on top of it. They are used by Banks for financial services provided for their customers.

The card issuer(banks)opens an account which is usually contributing a line of credit to the customer which the customer uses in-order to make payments. In spite of all these the credit card companies experience a lot of insecurities as in increase in card frauds with increase in new technologies.

Scammers take advantage of each and every loop holes and try to steal data. Data can be stolen using new technologies such as skimming and phishing.

Sometimes people get trapped in websites which are fake where they are asked to enter the security or personal details regarding their bank accounts or any other details. Usually, this information includes credit card number passwords of their online transactions or their credit card ccv. This is not the only means of fraud, scammers trap innocent customers by mails and otp extractions etc.

II. RELATED WORK

A. Credit card fraud

A fraudulent conduct is one in which a person intentionally or recklessly deprives someone of what they possess or money they are owed.

The unlawful use of a credit card or the theft of the owner's personal information is referred to as "Credit Card Fraud."

Two distinct types of credit card scams are linked by their different trick applications and actions. The first and second groups should be identified and described.

In order to commit app fraud, thieves either request for a new credit card from their bank or provide their existing card to a company that then uses it fraudulently. Duplicate fraud occurs when one person submits many applications using the same or similar descriptions (this is known as "duplicate fraud") (named identity fraud).

As a matter of fact, there are four basic categories of behavioral fraud: stolen/lost cards; mail theft; counterfeit cards; and current cardholder does not exist fraud. Credit card fraud happens when someone steals or loses a credit card.

Using a stolen credit card or the original card holder's details, a fraudster steals personal information from a bank and sends the stolen data to themselves. Descriptions of credit card and debit card fraud are not included. In the past, card information may be used to conduct remote conversations by mail, phone, or the internet. Fake cards are manufactured from card data in the second step (instead of the first).

B. Credit card Fraud detection

KNN, Artificial Intelligence, Random Forest, etc. are some of the current methods for detecting such fraud. In this approach that we have implemented we are using Logistic Regression Approach.

C. Detection Process

Step 1: Retrive a new transaction from the dataset that is to be classified as have it as a classified transaction.
Step 2: Take a transaction that was a non-fraud transaction from the dataset.
Step 3: Analyze the collection of legal transactions of various customers by using the Logistic Regression technique to determine the difference between the transactions.

II. EXPERIMENTAL SETUP AND METHODS

For the sake of this research, we'll be using datasets that fall within the purview of Logistic Regression. NumPy, Pandas, Keras, Scikit-Learn, and TensorFlow were some of the libraries used in the implementation. Rstudio was used to clean data on occasion. For the whole transactional picture, PowerBI is also integrated. Collection of data, preparation of data, analysis and training of the classifier algorithm and testing are some of the several processes involved in transactions. During the preprocessing step, the data is transformed into a usable format and sampled. PCA (Principal Component Analysis) is used to identify and reduce features in the dataset during the analysis stage. Classifier algorithms are built and supplied withprocessed data during the training phase. Using True Positive, False Positive, True Negative, and False Negative as evaluative criteria, we conduct tests to determine whether or not the transactions are beneficial overall. Based on these classifiers, the performance comparison is evaluated in terms of accuracy, sensitivity, specificity, and precision [36]-[58].

A. Dataset

PCA characteristics v1-v28, Time, Amount, and Class 0 and 1 are included in the dataset because of confidentiality concerns, which includes transactions of European credit card users for two days in September 2013. Only 0.17% of all transactions are in the "good" category. All transactions and the fraud transactions log file were combined to form this dataset. Unbalanced and heavily weighted in favour of the positive class. The selection of PCA characteristics resulted in the development of a total of 28 significant components. This study made use of a total of thirty different qualities as inputs. Transactions are organised into groups according on the amount of time that has passed since the beginning of the dataset. The 'amount' attribute of a transaction is used to reflect the monetary value of the transaction. The 'class' feature is the target class in the binary classification system, and it takes the value 1 (the positive case) as fraud and the value 0 (the negative case) as non-fraud.

B. Data Cleaning

When it comes to data cleansing, it's critical to fill up any gaps. You can solve the problem by disregarding the whole set, but most of these solutions are likely to have an impact on the data. In addition, the date time column was split into two after adjustments such as the removal of pointless columns.

C. Data Integration

It was necessary to merge both the bogus and real records into a single file before any further changes could be made.

D. Data Transformation

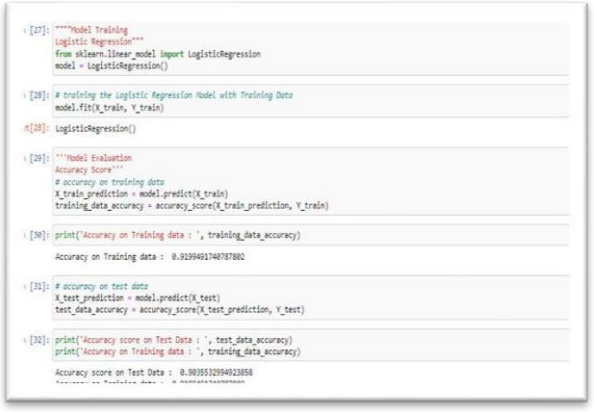
All of the category's information was condensed into a single number value for easy reference. Multiple data kinds and ranges are included in the transaction dataset. As a result, data normalization is a component of data transformation.

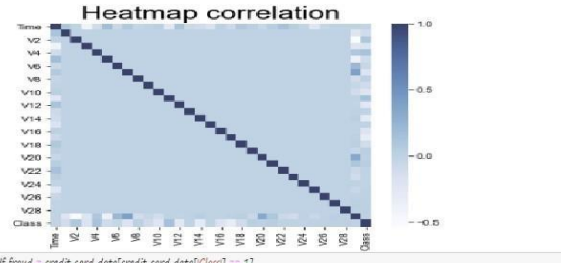
E. Data Reduction

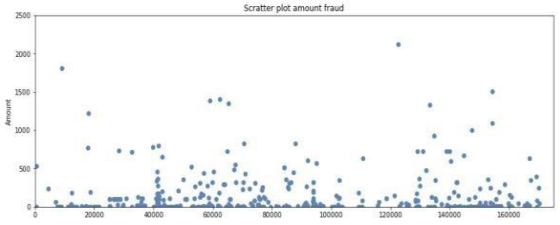
Dimensional reduction is the method used here. In the field of data transformation, PCA, or principal component analysis, is a well-known and widely used approach. The problem of feature selection is resolved by using this technique from a numerical analysis standpoint. Using PCA, the optimal number of principle components was identified, allowing the process of feature selection to go smoothly.

F. Logistic Regression

Classification jobs mostly employ this technique; however, it may also be used for regression. A binary corresponding to one of the classes is the result of a Logistic Regression model. Predicting categorical variables with the use of dependent variables is a common application of this technique. Probability values ranging from 0 to 1 may be calculated with ease using this approach (1)



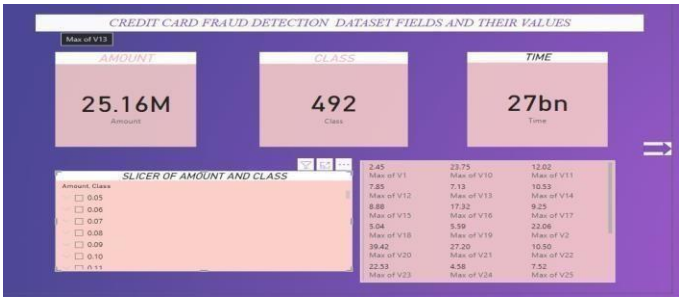


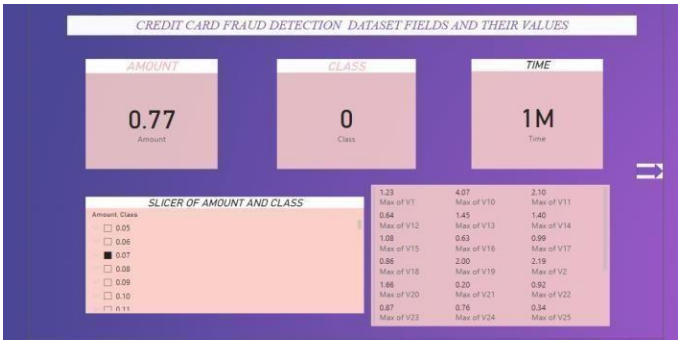


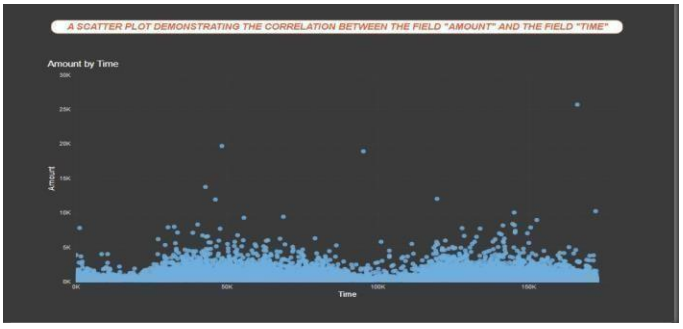
The x-values of the dataset's instances are transformed into a range of 0 to 1 using the logistic function. If the sigmoid's value is more than 0.5, it is deemed to be 1; otherwise, it is 0. The gradient ascent is then computed for each feature value in the dataset. With such big datasets, it is more efficient to employ gradient ascent, which changes weights with a single instance at a time.

G. Data Visualisation using Powerbi

The system also involves PowerBi visualization that gives you live visualization of the analysis based on the requirements.







CONCLUSION

To preserve the safety of cardholders' personal information and to guard against credit card fraud, the introduction outlined several methods for detecting and stopping fraudulent transactions. Credit card issuers should not depend only on one method of acquiring new customers. Even yet, they must use many methods at the same time in order to guarantee the safety of their client's private information and money. Any credit card firm may benefit from using these fraud detection strategies, which can reduce yearly losses due to credit card theft. Their yearly profit also grows as a result. The most cost-effective and time-efficient method should be used to the systems of companies. Using a random credit card number that generates a unique 16-digit credit card number for each online transaction with a predetermined amount of money is an option for future development. The cardholder's and the card issuer's security may be enhanced by combining this strategy with additional anti-fraud measures. We can still acquire the findings using logistic regression, which we will try in the future if we increasethe accuracy and reduce the number of fraudulent transactions.

ACKNOWLEDGEMENTS

We wish to acknowledge Dr. M. RamKumar for his effort in the experimentation carried out and the team for the source and description of the credit card fraud data.

REFERENCES

[1] Abhinav Srivastava, Amal Kundu, Shamik sural, Arun Majumdar - Credit Card Fraud Detection Using Hidden Markov Model IEEE 2008

[2] Abrar Nadim , Ibrahim Mohammad Sayem , Aapan Mutsuddy ,Mohammad Sanaullah Chowdhury -Analysis of Machine Learning Techniques for Credit Card Fraud Detection IEEE 2019

[3] CLIFTON PHUA1, VINCENT LEE1, KATE SMITH1 & ROSS GAYLER2 A Comprehensive Survey of Data Mining-based Fraud Detection Research published by School of Business Systems, Faculty of Information Technology, Monash University, Wellington Road,Clayton, Victoria 3800, Australia

[4] John Richard D. Kho, Larry A. Vea -Credit Card Fraud Detection Based on Transaction Behaviour - published by Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017

[5] Nikita Shirodkar , Pratiksha mandrekar,Rohit Shet Mandrekar , Rahul Sakhalkar, K.M. Chaman Kumar , Shailendra Aswale - Credit Card Fraud Detection Techniques IEEE 2020

[6] Suman-Survey Paper on Credit Card Fraud Detection , Research Scholar, GJUS&T Hisar HCE, Sonepat published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2014

[7] S. H. Projects and W. Lovo, ―JMU Scholarly Commons Detecting credit card fraud : An analysis of fraud detection techniques,‖ 2020.

[8] N. R. Deepak and S. Balaji, "Performance analysis of MIMO-based transmission techniques for image quality in 4G wireless network," 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2015, pp. 1-5, doi: 10.1109/ICCIC.2015.7435774.

[9] Loganathan, R., & Kumaraswamy, Y. S. (2013). Active contour based medical image segmentation and compression using biorthogonal wavelet and embedded zerotree. Indian Journal of Science and Technology, 6(4), 4390-4395.

[10] Jotheeswaran, J., Loganathan, R., & Madhu Sudhanan, B. (2012). Feature reduction using principal component analysis for opinion mining. International Journal of Computer Science and Telecommunications, 3(5), 118-121.

[11] Loganathan, R., & Kumaraswamy, Y. S. (2011, December). An improved active contour medical image compression technique with lossless region of interest. In 3rd International conference on trendz in information sciences & computing (TISC2011) (pp. 128-132). IEEE.

[12] Loganathan, R., & Kumaraswamy, Y. S. (2010). Medical image compression using biorthogonal spline wavelet with different decomposition. IJCSE International Journal on Computer Science and Engineering, 2(9), 3003-3006.

[13] Loganathan, R., & Kumaraswamy, D. Y. (2012). Medical Image Compression with Lossless Region of Interest Using Fuzzy Adaptive Active Contour. In International Conference on Computational

Techniques and Mobile Computing (ICCTMC'2012) December (pp. 14-15).

[14] Loganathan, R., & Kumaraswamy, Y. S. (2002). Performance Evaluation of Image Compression for Medical Image. International Journal of Advanced Research in Computer Science and Software Engineering [2013] Vol, 4.

[15] Kurian, S., & Ramasamy, L. (2021). Securing Service Discovery from Denial of Service Attack in Mobile Ad Hoc Network (MANET). International Journal of Computer Networks and Applications, 8(5), 619-633.

[16] Khan, Z., & Loganathan, R. (2020, October). AutoLiv: Automated Liver Tumor Segmentation in CT Images. In 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE) (pp. 151-156). IEEE.

[17] Loganathan, R., Khan, F. A., Gulzar, I., Parray, I. N., & Bhat, F. A. (2020).A Survey on Prober: An automated network vulnerability scanner. International Scientific Journal of Contemporary Research in Engineering Science and Management, 5(2), 85-88.

[18] Loganathan, R., Aliya, B. B., Rehman, S. S. U., & Pasha, A. (2020). A Survey on Paperless Examination. International Scientific Journal of Contemporary Research in Engineering Science and Management, 5(2), 80-84.

[19] Khan, Z. (2020). Radiomics in Prostate MRI: A Review on Opportunities & Challenges. International Scientific Journal of Contemporary Research in Engineering Science and Management, 5(1), 7-10.

[20] Kurian, S., & Ramasamy, L. (2021). Novel AODV based service discovery protocol for MANETS. Wireless Networks, 27(4), 2497-2508.

[21] Patan, R., & Gandomi, A. H. (2021). Improved salient object detection using hybrid Convolution Recurrent Neural Network. Expert Systems with Applications, 166, 114064.

[22] Yuvaraj, N., Srihari, K., Dhiman, G., Somasundaram, K., Sharma, A., Rajeskannan, S., ... & Masud, M. (2021). Nature-InspiredBased Approach for Automated Cyberbullying Classification on Multimedia Social Networking. Mathematical Problems in Engineering, 2021.

[23] Natarajan, Y., Kannan, S., & Mohanty, S. N. (2021). Survey of Various Statistical Numerical and Machine Learning Ontological Models on Infectious Disease Ontology. Data Analytics in Bioinformatics: A Machine Learning Perspective, 431-442.

[24] Raja, R. A., Yuvaraj, N., & Kousik, N. V. (2021). Analyses on Artificial Intelligence Framework to Detect Crime Pattern. Intelligent Data Analytics for Terror Threat Prediction: Architectures, Methodologies, Techniques and Applications, 119-132.

[25] Kannan, S., Dhiman, G., Natarajan, Y., Sharma, A., Mohanty, S. N., Soni, M., ... & Gheisari, M. (2021). Ubiquitous Vehicular AdHoc Network Computing Using Deep Neural Network with IoT-Based Bat Agents for Traffic Management. Electronics, 10(7), 785.

[26] Yuvaraj, N., Raja, R. A., Karthikeyan, T., & Kousik, N. V. (2020). 11 Improved Privacy Preservation Framework for Cloud-Based Internet of Things. Internet of Things: Integration and Security Challenges, 165.

[27] Yuvaraj, N., Karthikeyan, T., & Praghash, K. (2021). An improved task allocation scheme in serverless computing using gray wolf Optimization (GWO) based reinforcement learning (RIL) approach. Wireless Personal Communications, 117(3), 2403-2421.

[28] Mariappan, L. T., & Yuvaraj, N. (2020). Analysis On Cardiovascular Disease Classification Using Machine Learning Framework. Solid State Technology, 63(6), 10374-10383.

[29] Karthick, S., Yuvaraj, N., Rajakumari, P. A., & Raja, R. A. (2021). Ensemble Similarity Clustering Frame work for Categorical Dataset Clustering Using Swarm Intelligence. In Intelligent Computing and Applications (pp. 549-557). Springer, Singapore.

[30] Yuvaraj, N., Raja, R. A., & Kousik, N. V. (2021). Privacy Preservation Between Privacy and Utility Using ECC-based PSO Algorithm. In Intelligent Computing and Applications (pp. 567-573). Springer, Singapore.

[31] Yuvaraj, N., Raja, R. A., Palanivel, P., & Kousik, N. V. (2020, April). EDM Process by Using Copper Electrode with INCONEL 625 Material. In IOP Conference Series: Materials Science and Engineering (Vol. 811, No. 1, p. 012011). IOP Publishing.

[32] Veerappan Kousik, N. G., Natarajan, Y., Suresh, K., Patan, R., & Gandomi, A. H. (2020). Improving Power and Resource Management in Heterogeneous Downlink OFDMA Networks. Information, 11(4), 203.

[33] Natarajan, Y., Raja, R. A., Kousik, D. N., & Johri, P. (2020). Improved Energy Efficient Wireless Sensor Networks Using Multicast Particle Swarm Optimization. Available at SSRN 3555764.

[34] Khadidos, A., Khadidos, A. O., Kannan, S., Natarajan, Y., Mohanty, S. N., & Tsaramirsis, G. (2020). Analysis of COVID-19 Infections on a CT Image Using DeepSense Model. Frontiers in Public Health, 8. 20.

[35] Yuvaraj, N., Srihari, K., Chandragandhi, S., Raja, R. A., Dhiman, G., & Kaur, A. (2021). Analysis of protein-ligand interactions of SARS-Cov-2 against selective drug using deep neural networks. Big Data Mining and Analytics, 4(2), 76-83.

[36] Yuvaraj, N., Karthikeyan, T., Praghash, K., & Reddy, K. H. (2021). Binary flower pollination (BFP) approach to handle the dynamic networking conditions to deliver uninterrupted connectivity. *Wireless Personal Communications*, *121*(4), 3383-3402.

[37] Maheshwari, V., Mahmood, M. R., Sravanthi, S., Arivazhagan, N., ParimalaGandhi, A., Srihari, K., ... & Sundramurthy, V. P. (2021). Nanotechnology-Based Sensitive Biosensors for COVID-19 Prediction Using Fuzzy Logic Control. *Journal of Nanomaterials*, *2021*.

[38] Natarajan, Y., Kannan, S., Selvaraj, C., & Mohanty, S. N. (2021). FORECASTING ENERGY GENERATION IN LARGE PHOTOVOLTAIC PLANTS USING RADIAL BELIEF NEURAL NETWORK. *Sustainable Computing: Informatics and Systems*, 100578.

[39] Natarajan, Y., Raja, R. A., Kousik, N. V., & Saravanan, M. (2021). A review of various reversible embedding mechanisms. *International Journal of Intelligence and Sustainable Computing*, *1*(3), 233-266.

[40] Kousik, N. V., Sivaram, M., Yuvaraj, N., & Mahaveerakannan, R. (2021). Improved Density-Based Learning to Cluster for User Web Log in Data Mining. In *Inventive Computation and Information Technologies* (pp. 813-830). Springer, Singapore.

[41] Yuvaraj, N., Praghash, K., & Karthikeyan, T. (2021). Data Privacy Preservation and Trade-off Balance Between Privacy and Utility Using Deep Adaptive Clustering and Elliptic Curve Digital Signature Algorithm. *Wireless Personal Communications*, 1-16.

[42] Arivazhagan, N., Somasundaram, K., Vijendra Babu, D., Gomathy Nayagam, M., Bommi, R. M., Mohammad, G. B., ... & Prabhu Sundramurthy, V. (2022). Cloud-Internet of Health Things (IOHT) Task Scheduling Using Hybrid Moth Flame Optimization with Deep Neural Network Algorithm for E Healthcare Systems. *Scientific Programming*, *2022*.

[43] Gobinathan, B., Mukunthan, M. A., Surendran, S., Somasundaram, K., Moeed, S. A., Niranjan, P., ... & Sundramurthy, V. P. (2021). A Novel Method to Solve Real Time Security Issues in Software Industry Using Advanced Cryptographic Techniques. *Scientific Programming*, *2021*.

[44] Yuvaraj, N., Raja, R. A., Karthikeyan, T., & Praghash, K. (2021). Improved Authentication in Secured Multicast Wireless Sensor Network (MWSN) Using Opposition Frog Leaping Algorithm to Resist Man-in-Middle Attack. *Wireless Personal Communications*, 1-17.

[45] Yuvaraj, N., Praghash, K., Raja, R. A., & Karthikeyan, T. (2021). An Investigation of Garbage Disposal Electric Vehicles (GDEVs) Integrated with Deep Neural Networking (DNN) and Intelligent Transportation System (ITS) in Smart City Management System (SCMS). *Wireless Personal Communications*, 1-20.

[46] Kumar, A. S., Jule, L. T., Ramaswamy, K., Sountharrajan, S., Yuuvaraj, N., & Gandomi, A. H. (2021). Analysis of false data detection rate in generative adversarial networks using recurrent neural network. In *Generative Adversarial Networks for Image-to-Image Translation* (pp. 289-312). Academic Press.

[47] Sangeetha, S. B., Sabitha, R., Dhiyanesh, B., Kiruthiga, G., Yuvaraj, N., & Raja, R. A. (2022). Resource Management Framework Using Deep Neural Networks in Multi-Cloud Environment. In *Operationalizing Multi- Cloud Environments* (pp. 89-104). Springer, Cham.

[48] Gowrishankar, J., Kumar, P. S., Narmadha, T., & Yuvaraj, N. (2020). A Trust Based Protocol For Manets In Iot Environment., International Journal of Advanced Science and Technology 29 (7), 2770-2775.

[49] Karthick, S., Yuvaraj, N., Rajakumari, P. A., & Raja, R. A. (2021). Ensemble Similarity Clustering Frame work for Categorical Dataset Clustering Using Swarm Intelligence. In Intelligent Computing and Applications (pp. 549-557). Springer, Singapore.

[50] Yuvaraj, N., Raja, R. A., & Kousik, N. V. (2021). Privacy Preservation Between Privacy and Utility Using ECC-based PSO Algorithm. In Intelligent Computing and Applications (pp. 567-573). Springer, Singapore.

[51] Daniel, A., Kannan, B. B., Yuvaraj, N., & Kousik, N. V. (2021). Predicting Energy Demands Constructed on Ensemble of Classifiers. In Intelligent Computing and Applications (pp. 575-583). Springer, Singapore.

[52] Yuvaraj, N., Raja, R. A., Kousik, N. V., Johri, P., & Diván, M. J. (2020). Analysis on the prediction of central line-associated bloodstream infections (CLABSI) using deep neural network classification. In

Computational Intelligence and Its Applications in Healthcare (pp. 229 - 244). Academic Press.

[53] Sangeetha, S. B., Blessing, N. W., Yuvaraj, N., & Sneha, J. A. (2020). Improving the training pattern in back-propagation neural networks using holt-winters' seasonal method and gradient boosting model. In Applications of Machine Learning (pp. 189-198). Springer, Singapore.

[54] Natarajan, Y., Raja, R. A., Kousik, D. N., & Johri, P. (2020). Improved Energy Efficient Wireless Sensor Networks Using Multicast Particle Swarm Optimization. Available at SSRN 3555764.

[55] Yuvaraj, N., Kousik, N. V., Jayasri, S., Daniel, A., & Rajakumar, P. (2019). A survey on various load balancing algorithm to improve the task scheduling in cloud computing environment. J Adv Res Dyn Control Syst, 11(08), 2397-2406.

[56] Yuvaraj, N., Chang, V., Gobinathan, B., Pinagapani, A., Kannan, S., Dhiman, G., & Rajan, A. R. (2021). Automatic detection of cyberbullying using multi-feature based artificial intelligence with deep decision tree classification. Computers & Electrical Engineering, 92, 107186.

[57] Natarajan, Y., Kannan, S., & Dhiman, G. (2021). Task scheduling in cloud using aco. Recent Advances in Computer Science and Communications, 13, 1-6.

[58] A. Author, "Book style with paper title and editor," in *Title*, 1nd ed. vol. 1, C. Editor, Ed. City: Publisher, 2017, pp. 10–50.

[59] F. Author, H. Author, and I. Author, "Journal style," *Journal*, vol. 1, Jan. 1999, pp. 140–151 [*Conference*, 2016, pp. 300-307].

[60] F. Author and Z. Author, "Journal style," *Journal*, vol. 2, pp. 1–21, Jan. 2015.

[61] J. A. Author, "Periodical style," *Periodical*, vol. 1, no. 1, pp. 30–38, Jan. 1999. DOI: 01XYZ.

[62] K. A. Author, "Published Conference Proceedings style," in *Proc. Conference*, City, 2017, pp. 18–26.

[63] L. A. Author and M. Author, "Presented Conference Paper style," presented at Meeting (Conference), City, Jan. 2–7, 2015.

[64] A. Author, "Conferences," Publisher, vol. 1, no. 1, pp. 10–18, Feb. 2010. DOI: 08XYZ.

D. C. Author, "Write this article," *Periodical*, vol. 1, no. 1,pp. 5-16, 2017. DOI: 78XYZ.