ML-BASED CREDIT CARD FRAUD DETECTION FOR ONLINE TRANSITION

Dr.ECCLESTON *Student, IT Department, BIT Durg*malakianurag@gmail.com

Dr.JULIE *Professor, IT Department, BIT Durg*jyothi.pillai@bitdurg.ac.in

Abstract: The use of a credit card without authorization by someone who is not the account owner is fraud in credit card transactions. The abuse may be stopped with the use of necessary preventative measures, and the behavior of such fraudulent acts can be researched to lessen it and safeguard against recurrence. In other terms, credit card fraud is the use of another person's credit card for personal gain when neither the cardholder nor the organization responsible for providing the card are aware that the card is being used. Monitoring user populations' behavior is a key component of fraud detection since it helps identify, detect, and prevent undesirable behaviors including fraud, intrusion, and defaulting. This is a really pertinent issue that has to be addressed by fields like machine learning and data science, where an automated solution is possible. From the standpoint of learning, this issue is particularly difficult since it is characterized by many characteristics, such class imbalance. There are much more legitimate transactions than fraudulent ones. Additionally, the statistical characteristics of the transaction patterns frequently vary over time. In this paper, we provide a method for identifying credit card fraud using machine learning.

Keywords: Creditcard,frauddetection, machinelearning, python

I. INTRODUCTION:

The use of a credit card without authorization by someone who is not the account owner is fraud in credit card transactions. The abuse may be stopped with the use of necessary preventative measures, and the behavior of such fraudulent acts can be researched to lessen it and safeguard against recurrence. In other terms, credit card fraud is the use of another person's credit card for personal gain when neither the cardholder nor the organization responsible for providing the card are aware that the card is being used. Monitoring user populations' behavior is a key component of fraud detection since it helps identify, detect, and prevent undesirable behaviors including fraud, intrusion, and defaulting. This is a really pertinent issue that has to be addressed by fields like machine learning and data science, where an automated solution is possible. From the standpoint of learning, this issue is particularly difficult since it is characterized by many characteristics, such class imbalance. There are much more legitimate transactions than fraudulent ones. Additionally, the statistical characteristics of the transaction patterns frequently vary over time. The deployment of a fraud detection system in the real world is not without difficulties, though. In instances from the real world,

the enormous volume of payment requests is swiftly reviewed by automated technologies to choose which transactions to allow. ones. Algorithms for machine learning are used to analyses all permitted transactions and flag any dubious ones. Thesereportsare investigatedbyprofessionalswho contactthecardholders to confirm if the transaction was genuine orfraudulent. The investigators provide a feedback to theautomated system which is used to train and update thealgorithmto eventuallyimprovethefraud detectionperformance over time. A payment card, such as a credit card or debit card, is used to perpetrate fraud, and this practice is referred to as credit card fraud. The goal might be to pay into another account that is under criminal control or to receive goods or services. To assist financial institutions in securely processing card payments and lowering card fraud, the Payment Card Industry Data Security Standard (PCI DSS) was developed.

II. COMPARISON FRAUD DETECTION APPROACHES

Work	Technique used	Dataset used	Pre- Processing	Performance Metrics	Result
[16]	Bayesian Network Classifier (HHEA), instance reweighing and probability threshold analysis.	PagSeguro (Brazilian Online Payment Service)	✓	HM between Precision and Recall, and Economic Efficiency.	Fraud BNC following Probability threshold - more beneficial than NB, TAN, SVM, Decision Trees, Logistic Regression
[11]	Genetic Algorithm and Scatter Search.	Major Bank of Turkey (industrial partner).	✓	Misclassification cost based on TP, TN, FP, FN, TFL, S, r., No. of frauds, Ratio of legitimate transactions, Class Imbalance	Improved existing performance by 200%.
[1]	Hidden Markov Model	NA	✓	TP, FP	80% accurate
[8]	Bayesian Belief, ANN	Provided by Serge Waterschoot at Europay International	✓	TP, FP	Bayesian Belief, better than ANN. 8% more frauds detected. But ANN detects faster.
[9]	Bagging Ensemble Classifier	Real world credit card dataset obtained from USCO-FICO competition.	✓	Fraud Catching Rate, False Alarm Rate, Balanced Classification Rate, Mathews Correlation coefficient.	Stable, Fraud catching rate is high, Independent of rate of frauds, suitable for highly imbalanced dataset.
[14]	Big Data Analytical Framework with Hadoop	German Dataset	✓	FP, TP.	Random Forest Decision Tree performs best in terms of accuracy and precision among LR, DT & DTRF.

III. METHODOLOGY:

Machine learning (ML) is the study of computer algorithms that learn from experience and make informed decisions by utilising data. [1] It is considered to be a product of human

thought. Without being specifically programmed to do so, AI computations build a model based on example data, known as "preparing information," to make predictions or decisions. AI calculations are used in a broad range of applications, such as medicine, email separation, speech recognition, and computer vision, where it is difficult or impossible to establish routine computations to carry out the necessary tasks. The artificial intelligence (AI) used in the arranging interface includes choice trees, random woods, fake neural networks, and dishonest Bayes. This AI computation will be compared to find the best accuracy outcomes.

A. PreprocessingData:

Preprocessing is used to separate, modify, scale, and standardize new features that will be used in the AI computation interaction. Preprocessing is used to transform unreliable data into reliable data. PCA (Guideline Segment Examination) is used for preprocessing in this inquiry, with the main focuses being extraction, change, standardization, and scaling. PCA is a linear change that is frequently employed in information pressure and is typically used to extract highlights from data that is scaled at a high level of dimension. PCA may simplify complicated data to more basic measures to reveal hidden components and improve information generation. PCA estimations contain covariance framework computations to restrict decline and magnify volatility.

B. DecisionTree:

Investigating extortion information with the use of decision trees can reveal hidden relationships between several anticipated information elements and an objective variable. Choice tree is great as the first step in the showing interaction in any case, when used as the final model of a few different procedures, since it connects misrepresentation information exploration and demonstrating. A choice tree may be used for arrangement calculations. It is a type of administered learning calculation. The dataset is divided into a few growing portions based on choice standards by using a choice tree. This choice principle is regulated by establishing a link between information and yield credits.

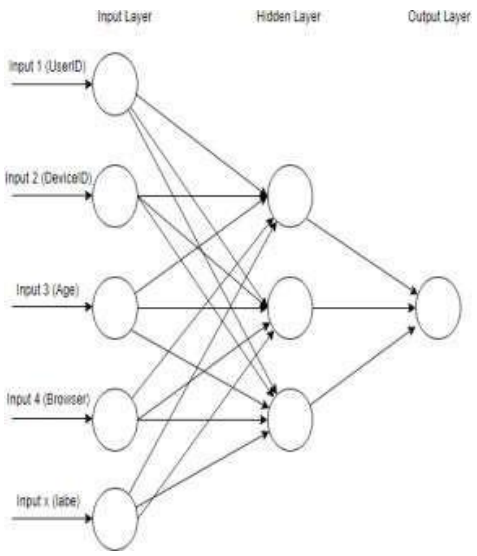
• **Root Hub:** This addresses the whole populace or test,andthisisadditionallypartitionedintoatleasttwo.Parting: This is the way toward partitioning a hub into atleasttwosub-hubs.

• **Decision Hub:** When a sub-hub is separated into a fewsubhubs.

• **Leaf/Terminal Hub**: Unknown hubsare called LeaforTerminalhubs.

• **Pruning:**Whenasub-hubiseliminatedfromachoice.

• **Branch/Sub-Tree:**Regionsofalltreesarecalledbranchesorsub-trees.



• **ParentandKidHub:**Ahub,whichisisolate dintosub-hubs

C. NeuralOrganization:

An organization or circuit of neurons is referred to as a neural organization. Using a cutting-edge viewpoint, a false neural organization is one that is constructed from artificial neurons or hubs. Accordingly, a neural organization is either a real neural structure made up of organic neurons or a fake neural organization used to solve artificial intelligence (AI) problems. The organic neuron's associations are shown as loads.
Apositiveweightmirrorsanexcitatoryassociati on,whilenegativequalitiesmeaninhibitoryasso ciations Each data source is given a weight adjustment and appended. A straight mix is a reference to this activity. Finally, an actuation

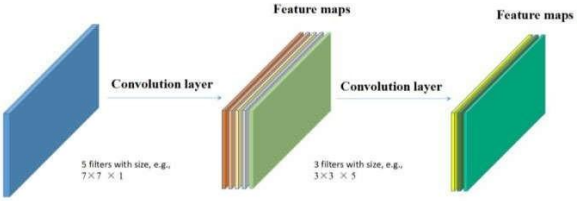
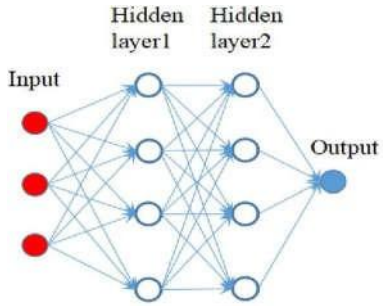


Figure2.1:ArchitectureofNeuralNetwork

work regulates the yield's sufficiency. For instance, a good range of yield
is often between 0 and 1, while it very well may be between 1 and 1. The computation. The aim behind the artificial intelligence approach known as neural organization is to apply a framework similar to that seen in Figure 2.1 to the human body, where hubs are connected to one another.



D. Deepconvolutionalneuralnetwork:

Onaccountofconvolution,thespecifictypeof boundary sharing causes the convolution layer to keep apropertycalledequivariancetointerpretatio n.Forinstance, for 2D pictures, convolution layer makes a 2Dguideofwherecertainhighlightsshowupi ntheinformation.Intheeventthatmovetheart iclesomewhatintheinfo,theelementportraya loftheinformationwilllikewisemove asimilarsum intheyield.

Figure2.2:Examplesoffullyconnectedneuralnet works

The red, white and blue circles are the input, hidden andoutput units for the neural network. The arrows indicatetheconnectionrelationshipbetween neuronunits.

Figure2.3:Examplesofconvolutionalnetwork swithaseriesoffilters

Convolutionalnetworkswithsmallerbyappl yingaseries of filters with size smallerthan the input size.Eachthincuboidisonefeaturechannelan disthe convolutionutput of one filter applied on the previousinput.

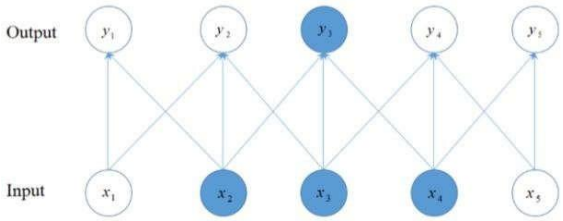


Figure2.4:Examplesofconvolutionoperations

The above row is obtained by convolution with kernelsize 3 applying to the bottom row. The arrows indicatewhich input units affect which output units. The bluecircles in the bottom row affect the output 3 and arecalled as the receptive field of y3. Other units in thebottom row do not have influence on the output units.Eachxiis theinputunitandyiistheoutputunit.

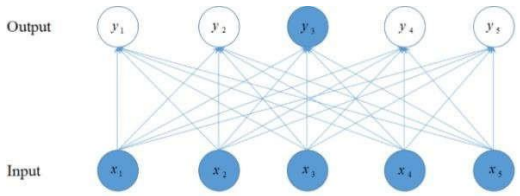


Figure2.5:Examplesoffullyconnectedneuralnet works

The above row is formed by matrix multiplication withfully connectivity. The arrows indicate which input unitsaffect which output units. All the units with blue circlesin the bottom row affect the output y3. Each xi is theinputunitandyiisthe outputunit.

IV. RESULT:

The most frequent issue is credit card fraud, which costs individuals and certain banks and credit card companies a lot of money. In order to prevent people from losing their wealth, as well as to benefit banked businesses, this project is working to create a model that more effectively distinguishes between transactions that are fraudulent and those that are not fraudulent by using the time and amount features in the Kaggel data set.

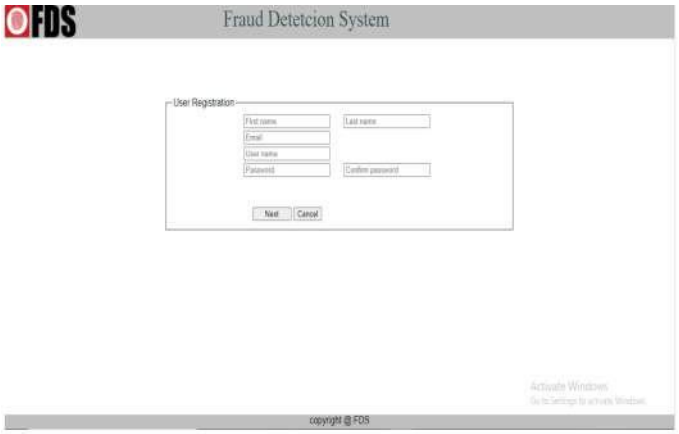


Figure 3.1: Login page GUI

Figure 3.2: Security question GUI

Figure 3.3: User home page GUI

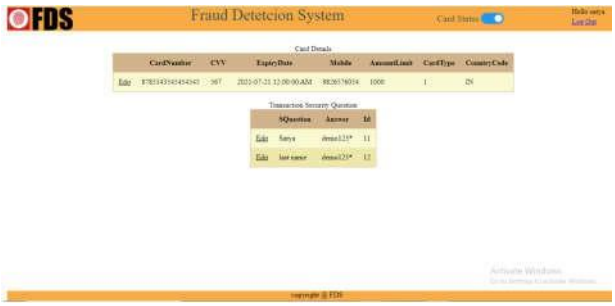


Figure 3.4: User home page with security question answers

Figure 3.5: Online shopping page

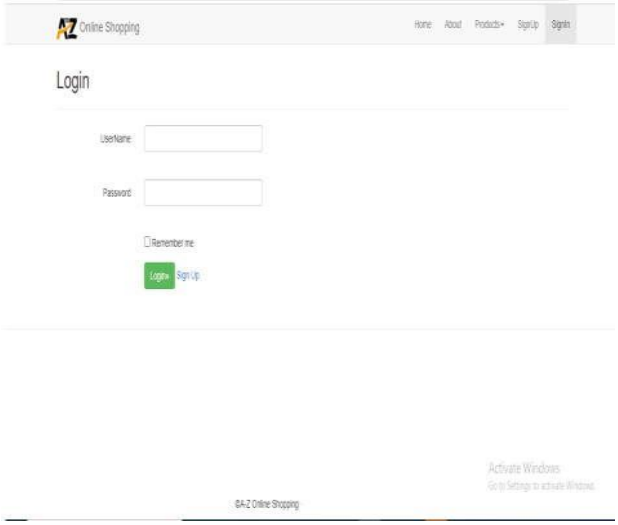


Figure 3.6: Online shopping page login

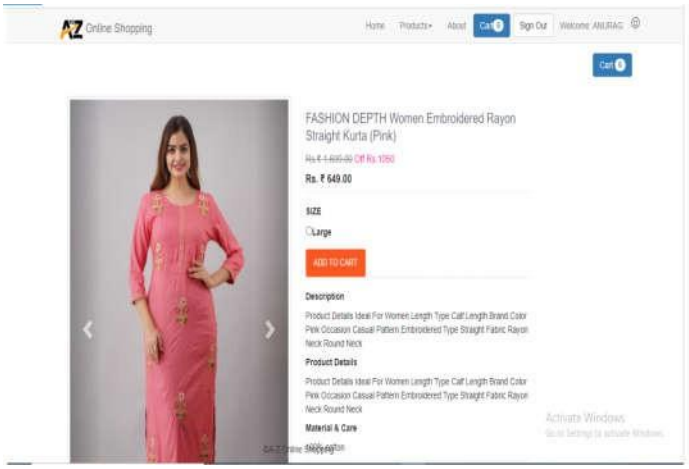


Figure 3.7: Selecting product

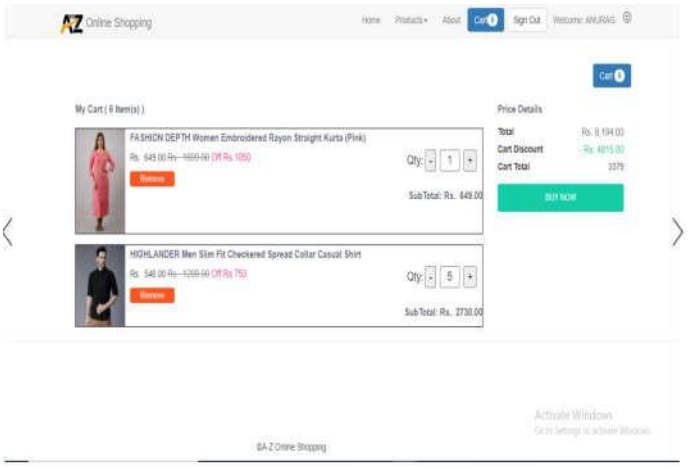


Figure 3.8: Selected products

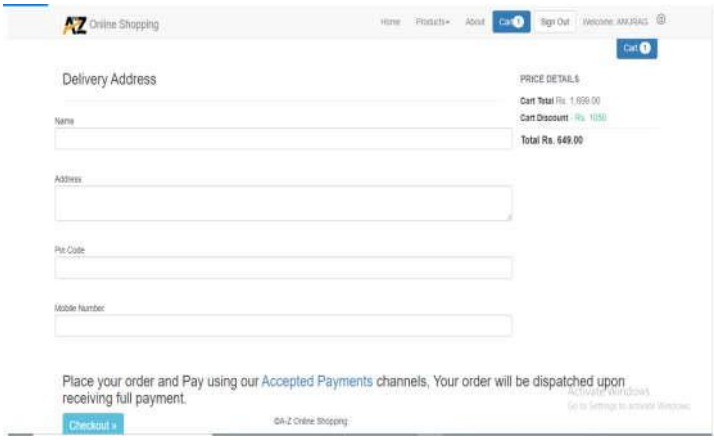


Figure 3.9: Adding delivery address

Figure 3.10: Payment gateway

Figure 3.11: Entering OTP GUI

Figure 3.12: Message display if transaction is more than the limit

V. CONCLUSION:

This system is capable of providing most of the essentialqualities needed to distinguish between fraudulent and legal transactions. It is harder to follow the behaviour and pattern of fraudulent transactions as technology develops. Although we have only discovered the fraudulent activity, we have not stopped it. Even if it is not simple, it is possible to stop known and undiscovered fraud in real time. The suggested architecture focuses on providing a fraud protection system to determine if a transaction is fraudulent or valid. It is primarily geared to detect credit card fraud in online payments. For the sake of implementation, it is assumed that the issuer and purchaser banks are affiliated. Exchange of best practises and promoting consumer awareness among individuals can be highly beneficial if this system is used in a real-world context to reducing the losses caused by fraudulent transactions. As technology advances, new checks can be added to the system to better understand the pattern of fraudulent transactions and to notify the relevant cardholders and bankers when fraud activity is discovered. This system can be further improved by making it secure through the use of certificates for both the merchant and the customer. For accurate fraud behavior identification, new data are required as the dataset used for day-to-day processing may grow stale. In order for the system to learn from both past and present data and be able to handle both, an incremental method is required. Fraudsters utilize a variety of fresh methods that are always evolving alongside new technology, making identification tough. Additionally, the type of access pattern may differ depending on the region (such as urban and rural areas), which might lead to a false positive detection. In this situation, it may be necessary to pay attention to new, numerous models with various access patterns in order to increase the efficacy. The security-related

problems restricting access to private data are resolved by privacy-preserving strategies used in dispersed environments.

REFERENCES:

[1]. AdiSaputra, Suharjito, "Fraud Detection using Machine Learning in e-Commerce", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 9, 2019.

[2]. BrankaStojanovi´c , Josip Boži´c, Katharina Hofer-Schmitz, Kai Nahrgang, Andreas Weber, Atta Badii, MaheshkumarSundaram, Elliot Jordan 3 and Joel Runevic, "Follow the Trail: Machine Learning for Fraud Detection in Fintech Applications", Sensors 2021, 21, 1594.

[3]. Elena-Adriana, Gabriela, "Light GBM Machine Learning Algorithm to Online Click Fraud Detection", IBIMA, 2019.

[4]. EvandroCaldeira, Gabriel Brandao, "Fraud Analysis and Prevention in e-Commerce Transactions", 2014 9th Latin American Web Congress, 978-1-4799-6953-1/14 $31.00 © 2014 IEEE DOI 10.1109/LAWeb.2014.23.

[5]. Gajendra Singh, Ravindra Gupta, Ashish Rastogi, Mahiraj D. S. Chandel, A. Riyaz, "A Machine Learning Approach for Detection of Fraud based on SVM", International Journal of Scientific Engineering and Technology (ISSN : 2277-1581) www.ijset.com, Volume No.1, Issue No.3, pg : 194-198.

[6]. Larisa Găbudeanu, Iulia Brici, Codrut,a Mare, Ioan Cosmin Mihai and Mircea Constantin S, cheau, "Privacy Intrusiveness in Financial-Banking Fraud Detection", Risks 2021, 9, 104.

[7]. S P Maniraj, Aditya Saini, Swarna Deep Sarkar, "Credit Card Fraud Detection using Machine Learning and Data Science", International Journal of Engineering Research, Volume 8 Issue 09, September-2019.

[8]. S. Venkata Suryanarayana, G. N. Balaji , G. Venkateswara Rao, "Machine Learning Approaches for Credit Card Fraud Detection", International Journal of Engineering & Technology, 7 (2) (2018) 917-920.

[9]. Suha M. Najem, Suhad M. Kadeem, "A Survey On Fraud Detection Techniques in E-Commerce", Techknowledge Journal, Volume 1, Issue 1, 2021.

[10]. Tuyls, B. Vanschoenwinkel, B. Manderick. "Credit Card Fraud Detection Using Bayesian and Neural Networks", BioData Mining, 2013.

[11]. S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and G. N. Surname, "Random forest for credit card fraud detection", IEEE 15th International Conference on Networking, Sensing and Control (ICNSC),2018.

[12]. Satvik Vats, Surya Kant Dubey, Naveen Kumar Pandey, "A Tool for Effective Detection of Fraud in Credit Card System", published in International Journal of Communication Network Security ISSN: 2231 – 1882, Volume-2, Issue-1, 2013.

[13]. Rinky D. Patel and Dheeraj Kumar Singh, "Credit Card Fraud Detection & Prevention of Fraud Using Genetic Algorithm", published by International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-6, January 2013.

[14]. M. HamdiOzcelik, EkremDuman, Mine Isik, TugbaCevik, "Improving a credit card fraud detection system using genetic algorithm", published by International conference on Networking and information technology, 2010.

[15]. Wen-Fang YU, Na Wang," Research on Credit Card Fraud Detection Model Based on

Distance Sum", published by IEEE International Joint Conference on Artificial Intelligence, 2009.

[16]. Andreas L. Prodromidis and Salvatore J. Stolfo; "Agent-Based Distributed Learning Applied to Fraud Detection"; Department of Computer Science- Columbia University; 2000.

[17]. Salvatore J. Stolfo, Wei Fan, Wenke Lee and Andreas L. Prodromidis; "Cost-based Modeling for Fraud and Intrusion Detection: Results from the JAM Project"; 0-7695-0490-6/99, 1999 IEEE.

[18]. Soltani, N., Akbari, M.K., SargolzaeiJavan, M., "A new user-based model for credit card fraud detection based on artificial immune system," Artificial Intelligence and Signal Processing (AISP), 2012 16th CSI International Symposium on., IEEE, pp. 029-033, 2012.

[19]. S. Ghosh and D. L. Reilly, "Credit card fraud detection with a neuralnetwork", Proceedings of the 27th Annual Conference on System Science, Volume 3: Information Systems: DSS/ Knowledge Based Systems, pages 621-630, 1994. IEEE Computer Society Press.

[20]. MasoumehZareapoor, Seeja.K.R, M.Afshar.Alam, "Analysis of Credit Card Fraud Detection Techniques: based on Certain Design Criteria", International Journal of Computer Applications (0975 – 8887) Volume 52– No.3, 2012.

[21]. Fraud Brief – AVS and CVM, Clear Commerce Corporation, 2003, http://www.clearcommerce.com.

[22]. Clear Commerce fraud prevention guide, Clear Commerce Corporation, 2002, http://www.clearcommerce.com

[23]. SamanehSorournejad, Zahra Zojaji , Reza EbrahimiAtani , Amir Hassan Monadjemi, "A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective ", IEEE 2016.

[24]. J. Cheng and R. Greiner, "Learning bayesian belief network classifiers: Algorithms and system," In Conferenceof the Canadian Society for Computational Studies of Intelligence, pp. 141-151, Springer, Berlin, Heidelberg,June-2011.

[25]. P. H. Swain and H. Hauska, "The decision tree classifier: Design and potential," IEEE Transactions on GeoscienceElectronics,15(3),142-147,1977.

[26]. C. J. Burges, "A tutorial on support vector machines for pattern recognition," Data mining and knowledgediscovery,2(2),121-167,1998.

[27]. R. S. Reshma, "Deep Learning Enabled Fraud Detection in Credit Card Transactions," International Journal ofResearchandScientificInnovation(IJRSI),July2018 .

[28]. Y. Pandey, "Credit card fraud detection using deep learning," International Journal of Advanced Research inComputerScience,May–June2017.

[29]. A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi and G. Bontempi, "Credit Card Fraud Detection: A RealisticModelingandaNovelLearningStrategy,"inIEEETransactionsonNeuralNetworksandLearningSystems,vol.29,no.8,pp.3784-3797,Aug.2018.

[30]. Z. Kazemi and H. Zarrabi, "Using deep networks for fraud detection in the credit card transactions," IEEE 4thInternationalConferenceonKnowledge-BasedEngineeringandInnovation(KBEI),Tehran,2017,pp.0630-0633.

[31]. Z.C.Lipton,J.Berkowitz,andC.Elkan,"Acriticalreviewofrecurrentneuralnetworksforsequencelearning,"arXivpreprintarXiv:1506.00019,2015.

[32]. X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," IEEE Transactions on Pattern Analysis &MachineIntelligence,(8),841-847,1991.

[33]. K.S.Shin,andY.J.Lee,"Ageneticalgorithmapplicationinbankruptcypredictionmodeling,"ExpertSystemswithApplications,23(3),321-328,2002.

[34]. T. K. Ho, "Random decision forests," In Proceedings of 3rd international conference on document analysis andrecognition,Vol.1,pp.278-282,IEEE,August-1995.

[35]. L.Delamaire,H.A.H.Abdou,andJPointon,"Creditcardfraudanddetectiontechniques:areview,"BanksandBanksystems,4(2),57-68,2009.

[36]. J.Kumar,A.K.Singh,A.Mohan,"Resource-efficientload-balancingframeworkforclouddatacenternetworks,"ETRIJournal,43(1),pp.53-63,2021.

[37]. D.Saxena,I.Gupta,J.Kumar,A.K.Singh,X.We

n,“ASecureandMultiobjectiveVirtualMachinePlacementFrameworkforCloudDataCenter,”IEEESystems Journal,2021.(ArticleinPress)DOI:10.1109/JSYST.2021.3092521

[38]. A. K. Singh, J. Kumar, “Secure and energy aware load balancing framework for cloud data centre networks,”ElectronicsLetters,55(9),pp.540-541,2019.

[39]. J. Kumar and A. K. Singh, “Cloud Resource Demand Prediction using Differential Evolution based Learning,”20197thInternationalConferenceonSmart Computing&Communications(ICSCC),2019,pp.1-5.

[40]. J. Kumar, A. K. Singh, “Dynamic resource scaling in cloud using neural network and black hole algorithm,”Proceedings on 5th International Conference on Eco-Friendly Computing and Communication Systems, ICECCS2016,pp.63-67.

[41]. J.Kumar,R.Goomer,A.K.Singh,“LongShortTermMemoryRecurrentNeuralNetwork(LSTM-RNN)BasedWorkloadForecastingModelforCloudDatacenters,”ProcediaComputerScience,125,pp.676-682,2018.

[42]. A. K. Singh, D. Saxena, J. Kumar, V. Gupta, “A Quantum Approach towards the Adaptive Prediction of CloudWorkloads,”IEEETransactionsonParalleland DistributedSystems,32(12),pp.2893-2905,2021.