iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: V Month of publication: May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51920

Fraud Detection in Credit Card Automated System using ML with AWS SageMaker

Samrajya Pujari [1], Kaveri Diwanji[2], Samarth Malegaonkar[3], Samreen Shaikh[4], Prof. Ashwini Bhosale[5]
[1, 2, 3, 4]Under-Graduate Student, [5]Assistant Professor Department of Computer Engineering, Genba Sopanrao Moze College Engineering, Balewadi, Pune

Abstract: The article discusses the need for the widespread adoption of payment systems, which has been driven by advances in technology. However, the issue of fraud remains a major concern for financial institutions, as there is no one-size-fits-all approach to detecting and preventing fraudulent transactions. Machine learning has been identified as a potential solution to this problem, but it requires the development of a reliable automated system capable of handling large volumes of data in real-time. In the article, the author details the structure and setup of an automated fraud detection system for payment systems that relies on a web service hosted on the cloud. The deployment of this system is justified by utilizing Amazon Web Services as the platform, which includes Amazon Fraud Detector and Amazon A2I task type to authenticate and validate forecasts that are deemed high-risk. One instance of developing a system for detecting anomalies on Amazon DynamoDB streams is presented by utilizing AWS SageMaker, AWS Glue, and AWS Lambda. The software product aims to prevent and detect fraud in payment systems, with a rapid detection time and integration with various business institutions. The article also highlights the importance of developing a specific methodology for implementing the software product for fraud detection in payment systems.

Keywords: Fraud detection, anomaly detection, machine learning, automated system, cloud computing, big data, data analysis

I. INTRODUCTION

The digital revolution of society has led to customers conducting more transactions autonomously. There is less need to visit physical financial institutions as payment systems and services become more prevalent. The rise of modern payment systems is being driven by the growing prevalence of online transactions and their incorporation into a variety of industries, such as financial institutions, online stores, logistics firms, insurance providers, and trading platforms. These payment systems offer a range of benefits, such as convenience, speed, simplicity, transparency, and greater control over financial transactions. However, the risks associated with such transactions are also increasing, specifically the risk of fraud. Consumers and businesses are increasingly faced with fraudulent transactions, resulting in the loss of ordered goods or services and depletion of funds from accounts. Financial losses and a decline in customer loyalty are the consequences for businesses, potentially leading to a loss of customers.

II. LITERATURE REVIEW

The utilization of data [1] has exposed various forms of fraudulent activities in the banking sector, such as social engineering, transfers through online banking, card-to-card transfers, mobile banking access interception, counterfeit mobile banking, embezzlement through SMS-banking, and purchases made via Apple Pay and Google Pay. Payment card fraud has been identified as the primary type of fraud in the banking industry. The study proposes different methods for combating fraud and highlights the benefits of modern technologies that employ fraud detection models and methods. The key objective of the research [2-3], [4-5] is to categorize models of fraudulent transactions using various techniques. The suggested course of action is to employ this approach in the detection of fraud instances that have already been observed. The research [6] suggests that transactions can be flagged as fraudulent if they differ from the usual behavior of the customer, based on the assumption that attackers' behavior is vastly different from that of account owners. The approach to risk management of fraud with payment systems, which involves a combination of the presented approaches, defined in the works [7-8]. Given the above, it is advisable to first develop and define a model of user behavior, and then detect fraud. Various methods and algorithms can be used to solve this problem. It is worth noting the study presented in [9] – the lack of an effective and accurate algorithm for fraud, which would be the standard for all financial transactions. Each technique has its advantages and disadvantages. In addition, approaches to fraud are dynamic and require constant revision of forecasts, and this is due to the fact that each business in which possible fraudulent schemes are unique and must rely on its own corporate system. An example of the application of unified models of methods under the conditions of specific institutions is presented in the paper [10].

From the standpoint of the application of algorithms for detecting fraudulent transactions in specific businesses, it is appropriate to apply a set of systematic methods of forecasting, detection and control of fraudulent transactions in financial systems, which in [11] were called the fraud detection technologies. Models, methods and algorithms of machine learning are the basis of such technology. The disadvantage of this work is that it does not present the adaptation of the developed model into a software product. However, the market presents a large number of automated systems aimed at preventing fraud, which are called anti-fraud systems [13]. The literature is dominated by information about comprehensive systems for detecting bank fraud. At the heart of such systems are analytical platforms that allow you to implement logic in individual segments. Typically, such systems are used in the banking sector. The examples of such systems are: ARIC White Label by Featurespace, FICO Application Fraud Manager by FICO, FRAUD-Analysis by BSS, IBM Safer Payments by IBM, Fraud and Security Intelligence (SAS FSI) [13-14].

The paper [15-16] presents the principles of operation of systems aimed at identifying instruments of banking fraud. Examples of such systems are: Digital Banking Fraud Detection, WebSafe, IBM Trusteer Rapport, ThreatMetrix, Group-IB Secure Bank, etc.

There are highly specialized systems for detecting signs of bank fraud such as FPS.Bio, SmartTracker.FRAUD.

As well, there are mixed systems for counteracting bank fraud: RSA Adaptive Authentication and Transaction Monitoring, BI.ZONE Cloud Fraud Prevention.

The use of such systems is usually associated with additional costs. In addition, the experience of using ready-made software products shows that such products are not aimed at the concept of detecting local business fraud based on specific available data for an individual company. The implementation of an automated fraud detection system is much broader than the choice of system, and should take into account the range of tasks from the implementation of ideology among employees, formalization of procedures for collecting and storing information to changes in organizational structure and distribution of team roles. This is to some extent reflected in the work [17]. But in the work, there is no description of the automated system. The specialized automated system does not aim to compete with the presented ones, but, on the contrary, will complement their functions.

III. FORMULATION OF THE PROBLEM

The objective of this work is to address the challenge that although machine learning models and algorithms already exist to detect fraud in payment systems, the current environment with its large data volumes, high performance requirements, and need for optimization necessitates not just building the model, but also implementing it effectively in the application environment. In essence, detecting fraud in payment systems is more than just creating the algorithmic core that can identify fraud based on input data, but also developing a reliable and robust automated system that can manage real-time data flows under high load and efficiently operate the algorithmic core, which is based on previously developed models. This system should be tailored to the specific needs of individual businesses, and it is of both practical and scientific interest.

IV. THE AIM OF STUDY

The aim of this project is to create an automated system that can analyze incoming transaction data in real-time and distinguish between regular and fraudulent transactions. The proposed system should take into account the specific characteristics of each company in detecting fraudulent transactions. The project involves several research tasks, including developing the software product lifecycle, identifying technical challenges in developing the automated system, and assessing the implementation of the system in a business setting. The end result of the project will be a prototype of the automated system, including its architecture, software infrastructure, and interfaces.

V. RESEARCH RESULTS

The task involves identifying and setting up the necessary infrastructure to construct an automated system for detecting fraudulent transactions in payment systems. The emergence and expansion of automated systems, the shift from ad-hoc automation to strategic development of corporate IT systems, and the use of design and business-focused technologies have paved the way for detecting fraudulent transactions.

However, it is acknowledged that developing automated technology for preventing fraud is challenging, and it involves addressing issues such as selecting appropriate models, implementing the model as automated technology, and integrating automated technology into the business's operational processes.

1) the model choosing;
2) deployment of the model in the form of automated technology;
3) introduction of automated technology in the practice of using business institutions.

A. Choice of Model

The authors have used their own research, which is detailed in the paper [12], to address this problem. The model was developed using various attributes of transactions, including step – displays the unit of time in reality. The transactions in the dataset were completed within 743 hours. That is, “1” – the first hour of observation, “743” – 743rd hour of observation; type – type of transaction (cash replenishment, cash withdrawal, transfer of funds to the account, payment for goods or services, money transfer); sum – transaction sum; nameOrig – the client who initiated the transaction; oldBalanceOrig – the client's initial balance before the transaction; newBalanceOrig – customer balance after the operation; nameDest – identifier (identifier) of the recipient of the transaction; oldBalanceDest – the initial balance of the recipient; newBalanceDest – the balance of the recipient after the operation; isFraud – indicates if the transaction is a fraud (1) or not a fraud (0).

Therefore, the study examined a dataset of banking transactions carried out by individuals independently, such as those performed using mobile banking, cards, or terminals, without the involvement of a bank or other regulatory authority. It is important to consider the variability of the dataset in both the models and the automated system.

B. Deploying The Model In The Form Of Automated Technology

A recommended approach is to integrate the fraud detection system into the overall automated structure of the company. To accomplish this, the authors propose creating a cloud-based web service dedicated to analyzing data and determining whether a transaction is suspicious or not. Cloud technologies offer benefits such as improved scalability, availability, and cost optimization, enabling businesses to quickly respond to challenges. By providing a user-friendly environment for developers and businesses to collaborate, development and operations specialists can increase efficiency and shorten the development cycle, resulting in faster product launches. This approach can lead to significant changes in the development of cloud technologies, facilitate more rapid responses to the needs of enterprises, and decrease the cost of processes such as testing and deployment [18].

The several advantages of implementing a cloud-based system for fraud detection, including the ability to rapidly scale resources and improve productivity, optimize costs, and reduce time to market. The next step was to choose a cloud platform, and after assessing the benefits, the Amazon Web Services (AWS) solution was found to be the most comprehensive and widely used in the world. The authors chose AWS for its ease of use and flexibility, as it allows for quick and secure hosting of both new and existing applications based on the Software as a Service (SaaS) model. AWS also offers a range of customizable services, including operating systems, programming languages, and databases, that can be easily migrated to the cloud environment. This makes it simpler to develop and deploy fraud detection software and preserves the ability to create new solutions. [19]

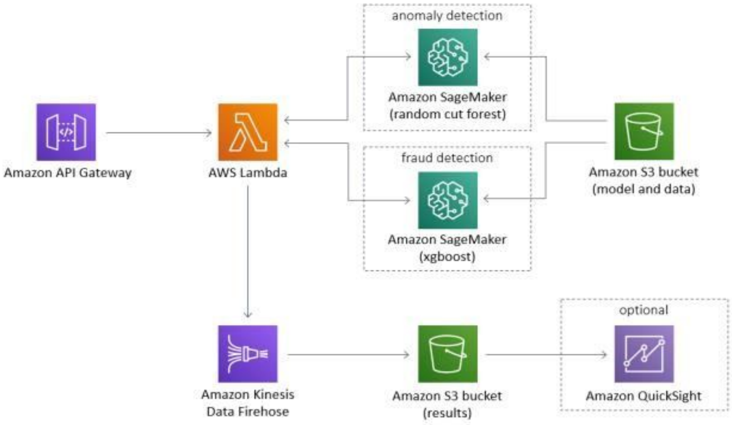

anomaly detection
Amazon SageMaker (random cut forest)
fraud detection
Amazon SageMaker (xgboost)
Amazon API Gateway
AWS Lambda
Amazon S3 bucket (model and data)
optional
Amazon Kinesis Data Firehose
Amazon S3 bucket (results)
Amazon QuickSight

Fig. 1. Basic architecture of the automated fraud detection system in payment systems (PTFD) based on [21]

The AWS platform was selected based on its evaluated benefits, such as its ease of use, allowing quick and secure hosting of both new and existing SaaS-based applications using the AWS management console or Web Services API with detailed documentation. It is also flexible, providing the ability to select operating systems, programming languages, internet application platforms, databases, and other necessary services, creating a virtual environment for software downloads for fraud detection, simplifying the migration process and preserving the ability to create new solutions. Moreover, it is cost-effective, as payment is only made for the computing power, storage capacity, and other resources used without prior commitments or long-term contracts. Lastly, it is reliable since it is the virtual basis for the multibillion-dollar Internet business Amazon.com, which has proven quality in practice.

The advantages of using AWS for implementing an automated fraud detection system include its scalability, high performance, security, cost-effectiveness, and reliability. AWS tools such as Auto Scaling and Elastic Load Balancing provide scalability, while AWS takes an integrated approach to security and infrastructure strengthening. The automated fraud detection system is designed to deploy a machine learning model and an example of a transaction data set to teach the model to recognize fraud patterns. This solution allows for the automation of detecting potentially fraudulent activity and sending it for verification. The system runs automated transaction processing on a data set or on a user's own data set. The solution includes the AWS CloudFormation template, which deploys an example of a set of credit card transactions, and an instance of Amazon SageMaker, which trains a controlled and unmanaged learning model. The system generates a continuous stream of transaction classification requests and uses Amazon Kinesis Data to deliver processed transactions to another segment of Amazon S3 for storage. Finally, the architecture employs Amazon Fraud Detector to perform low-delay fraud predictions and evaluate events on an hourly or daily schedule. The approach uses the Amazon S3 event message to run a lambda function that handles the CSV event file stored in Amazon S3 when the file is loaded into the S3 input segment. The function runs each event via Amazon Fraud Detector to generate predictions using a detector (ML model and rules) and loads the prediction results into the original S3 segment.

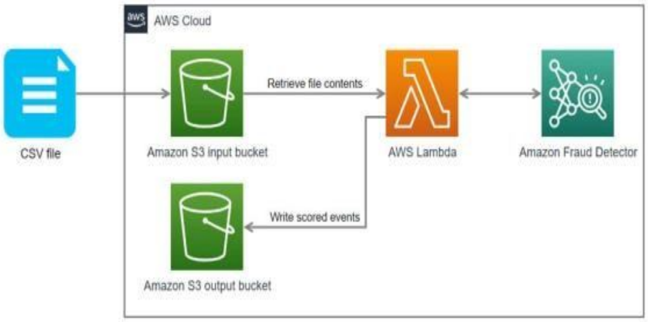


Fig. 2. Architecture of the automated fraud detection system with start of the classifier with an arbitrary interval [21]

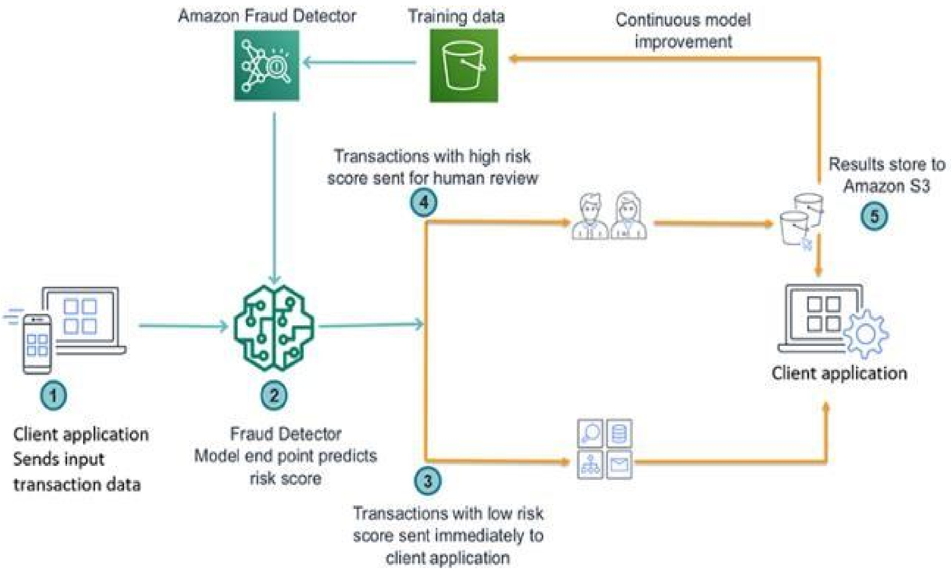

Amazon Fraud Detector
Training data
Continuous model improvement
Transactions with high risk score sent for human review
4
Results store to Amazon S3
5
Client application
1
Client application Sends input transaction data
2
Fraud Detector Model end point predicts risk score
3
Transactions with low risk score sent immediately to client application

Fig. 3. Real-time fraud detection automated system architecture and integration with client applications [21]

Additionally, you can enhance the system's performance by utilizing Amazon A2I, a machine learning service that streamlines the development of workflows involving the human validation of machine learning models. By leveraging Amazon A2I, you can delegate the tedious task of creating review systems or managing multiple reviewers, making the development process more efficient. The main idea of the solution is illustrated in the following architecture diagram [23].

The workflow consists of the following stages (Fig. 3.):

1) The client program sends information to the Amazon Fraud Detector endpoint.

2) Amazon Fraud Detector predicts a risk assessment (ranging from 0 to 1,000) for input using a historical learning machine learning model. A score of 0 means that there is no risk of fraud, and a score of 1,000 indicates that the risk of fraud is maximum.

3) If the risk assessment for a particular forecast falls below a predetermined threshold, no further action is taken.

4) If the risk assessment exceeds a predetermined threshold (for example, 900 points), the Amazon A2I cycle starts automatically and sends predictions for human verification to the Amazon A2I. Employees of the company can be private personnel. They open the Amazon A2I interface, review the case and make a decision (approve, reject or send it for further verification).

5) The result of approving or rejecting a private workforce is stored in the Amazon Simple Storage Service (Amazon S3). With Amazon S3 it can be sent directly to the client program.

The following steps are required to configure the solution:

a) Training and deployment of the model in Amazon Fraud Detector using historical data.

b) Configure the Amazon A2I cycle for staff using Amazon Fraud Detector.

c) Using the model to predict the risk assessment for given new input data.

d) Customize Amazon A2I workflow and cycles.

Also, to complete the process, consider a solution to detect fraud (anomalies) using the services Amazon DynamoDB Streams and Amazon SageMaker, which will complement the previous implementation and work consistently with it [24]. This solution has the following advantages:

- make better use of available resources to detect anomalies. For example, if you use Amazon DynamoDB streams for disaster recovery (DR) or for other purposes, you can use the data in that stream to detect anomalies. In addition, the backup storage is usually low. Low awareness data can be used for training data. automatic retraining of the model with new data on a regular basis, without user intervention.

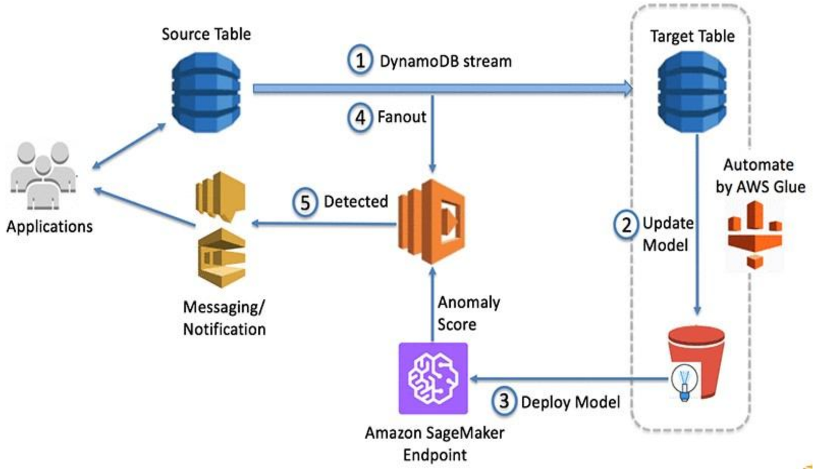


Fig. 4. Architecture for detecting anomalies in transaction data flows using the implemented automated system [21]

To simplify the use of Amazon SageMaker Random Cut Forest, a flexible distributed learning option that adapts to specific workflows in a secure and scalable environment, the data goes through several stages in the system's architecture (Fig. 4). First, changes are captured by DynamoDB and saved in the DynamoDB stream. AWS Glue retrieves data from the DynamoDB table and trains a model using Amazon SageMaker, creating or updating model artifacts on Amazon S3. The updated model is then used by AWS Glue to detect anomalies in real-time using the Random Forest classifier. The AWS Lambda function analyzes data from the DynamoDB stream, calls the Amazon SageMaker endpoint, and warns user programs when anomalies are detected. By combining all the previous subsystems, the fraud detection system (Fig. 5) will have two artifacts at the entrance to the system under development: an implemented algorithmic model for detecting fraud (with static code) and an input set of transaction data (dynamic and adaptable).

The system will be initialized and operated in three steps: Construction, Training, and Placement. Once these steps are successfully completed, an API access point will be created to communicate with the implemented automated system. To make this process even easier, Amazon A2I, a machine learning service, can be used to simplify the creation of workflows using the necessary machine learning models for human validation, thereby eliminating the need for manual review systems or large numbers of reviewers.

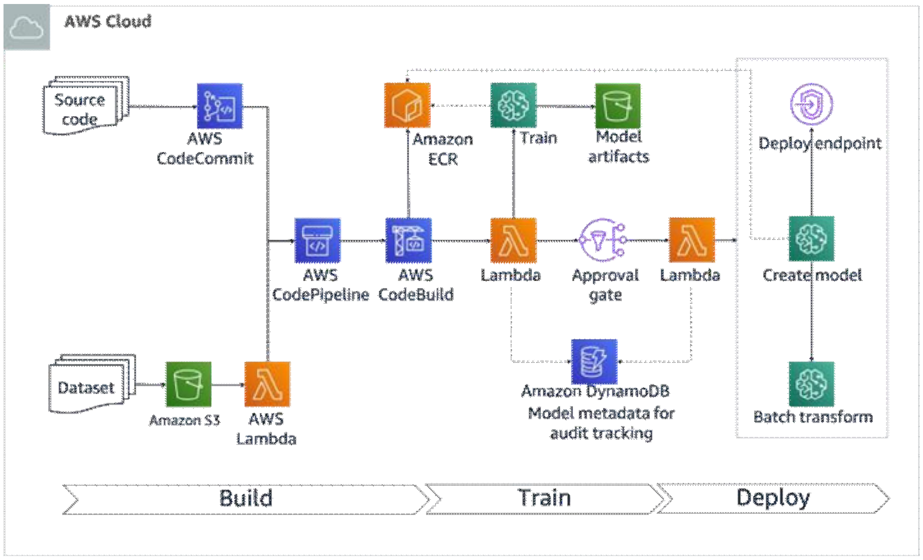


Fig. 5. Architecture of the full cycle automated fraud detection system [21]

The end result is a software product that addresses the problem of preventing and detecting fraud in payment systems. The solution is based on machine learning and is a fully managed service that includes all the necessary tools for creating, deploying, and managing fraud detection models. The model can be optimized with the use of business rules that allow for greater control over its operation and the deployment of its results in the form of easy-to-use APIs.

The system automates the complex stages of creating machine learning models, making it possible to detect fraud in just a matter of minutes. This is achieved without the need for any machine learning knowledge or programming skills. The system can be easily customized to meet the unique needs of any business scenario using patterns of fraudulent actions obtained through data analysis. This results in a high level of accuracy and reduces the number of false positives.

VI. CONCLUSIONS

The paper provides a detailed description of an automated fraud detection system, including its architecture, principles, and operational models, as well as the setup of its infrastructure. The use of Amazon Fraud Detector for online fraud detection and Amazon A2I for human fraud detection workflows is demonstrated, with the latter being highly customizable to confirm high-risk predictions. An example of building an anomaly detection system using Amazon DynamoDB streams, Amazon SageMaker, AWS Glue, and AWS Lambda is also presented, which can be adapted to specific use cases as AWS Glue is highly flexible and supports adding new data sources. The implementation of such a system for fraud detection in payment services is suggested to be treated as a project, and it is recommended to develop a specific methodology for its implementation in financial institutions, which could lead to further advancements in payment systems research.

REFERENCES

[1] Dubina, M. V., Sadchikova, I. V. & Seredyuk, I. O. “Conceptual approaches to increasing the level of security of the banking payment environment of Ukraine” (in Ukrainian). Available from: https://www.business-inform.net/export_pdf/business-inform-2020-3_0-pages-349_359.pdf

[2] Lebichot, B. & Le Borgne, Y.-A. “Deep-Learning Domain Adaptation Techniques for Credit Cards Fraud Detection”. In: Oneto, L., Navarin, N., Sperduti, A., Anguita, D. (eds.) Recent Advances in Big Data and Deep Learning. Publ. Springer. New York: 2019. p. 78–88.

[3] Caelen, O. & Smirnov, E. N. “Improving Card Fraud Detection through Suspicious Pattern Discovery”. In: Benferhat, S., Tabia, K., Ali, M. (eds.) Advances in Artificial Intelligence: From Theory to Practice. Publ. Springer. New York: 2017. p. 181–190.

[4] Pozzolo, A. D., Caelen, O., Bontempi, G. & Johnson, R. A. “Calibrating Probability with Undersampling for Unbalanced Classification”. Paper presented at the 2015 IEEE Symposium Series on Computational Intelligence.

[5] Lebichot, B., Le Borgne, Y. A., He-Guelton, L., Oblé, F. & Bontempi, G. “Deep-Learning Domain Adaptation Techniques for Credit Cards Fraud Detection”. In: Oneto L., Navarin N., Sperduti A., Anguita D. (eds). Recent Advances in Big Data and Deep Learning. INNSBDDL 2019. Proceedings of the International Neural Networks Society. Publ.

[6] Sorournejad, S. Z. Zojaji, R. E. & Atani Hassan Amir. “Monadjemi Fraud Detection Techniques”. Data and Technique Oriented Perspective. Cornel University Library. 2016. – Available from: https://arxiv.org/ftp/arxiv/papers/1611/1611.06439.pdf

[7] Kuznietsova, N. V. “Analysis and forecasting the risks of credit card fraud”. Informatics and Mathematical Methods in Simulation (in Ukrainian).

[8] Wang Hongbin. “Research and Application of web Log Mining Technology Based on Distributed Computing Platform [D]”

[9] “Cloud computing”. SoftServe. 2021. – Available from: https://www.softserveinc.com/ukua/services/cloud-devops.

[10] “Overview of Amazon Web Services”. 2021. – Available from: https://d1.awsstatic.com/ whitepapers/aws-overview.pdf.

[11] “Cloud Computing Solutions Architect: A Hands-On Approach”. A Competency-based Textbook for Universities and a Guide for AWS Cloud Certification and Beyond by Arshdeep Bahga, Vijay Madisetti. VPT.

[12] “AWS Documentation”. 2021. – Available from: https://docs.aws.amazon.com/ index.html?nc2=h_mo.

[13] “AWS Certified Cloud Practitioner Study Guide: CLF-C01 Exam”. 1st Edition by Ben Piper, David Clinton.

[14] “Architecting Cloud Computing Solutions: Build cloud strategies that align technology and economics while effectively managing risk”, by Kevin L. Jackson, Scott Goessling

[15] Mezentseva, O. O. & Kolomiiets, A. S. “Optimization of Analysis and Minimization of Information Losses in Text Mining”. Herald of Advanced Information Technology. Publ. Science i Technical. Odesa: Ukraine. 2020; Vol.3 No.1: 373–382. DOI:10.15276/hait.

[16] A. K. Singh, "Detection of Credit Card Fraud using Machine Learning Algorithms," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 673-677, doi: 10.1109/SMART55829.2022.10047099.

[17] A. Biswas, R. S. Deol, B. K. Jha, G. Jakka, M. R. Suguna and B. I. Thomson, "Automated Banking Fraud Detection for Identification and Restriction of Unauthorised Access in Financial Sector," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 809-814, doi: 10.1109/ICOSEC54921.2022.9951931.

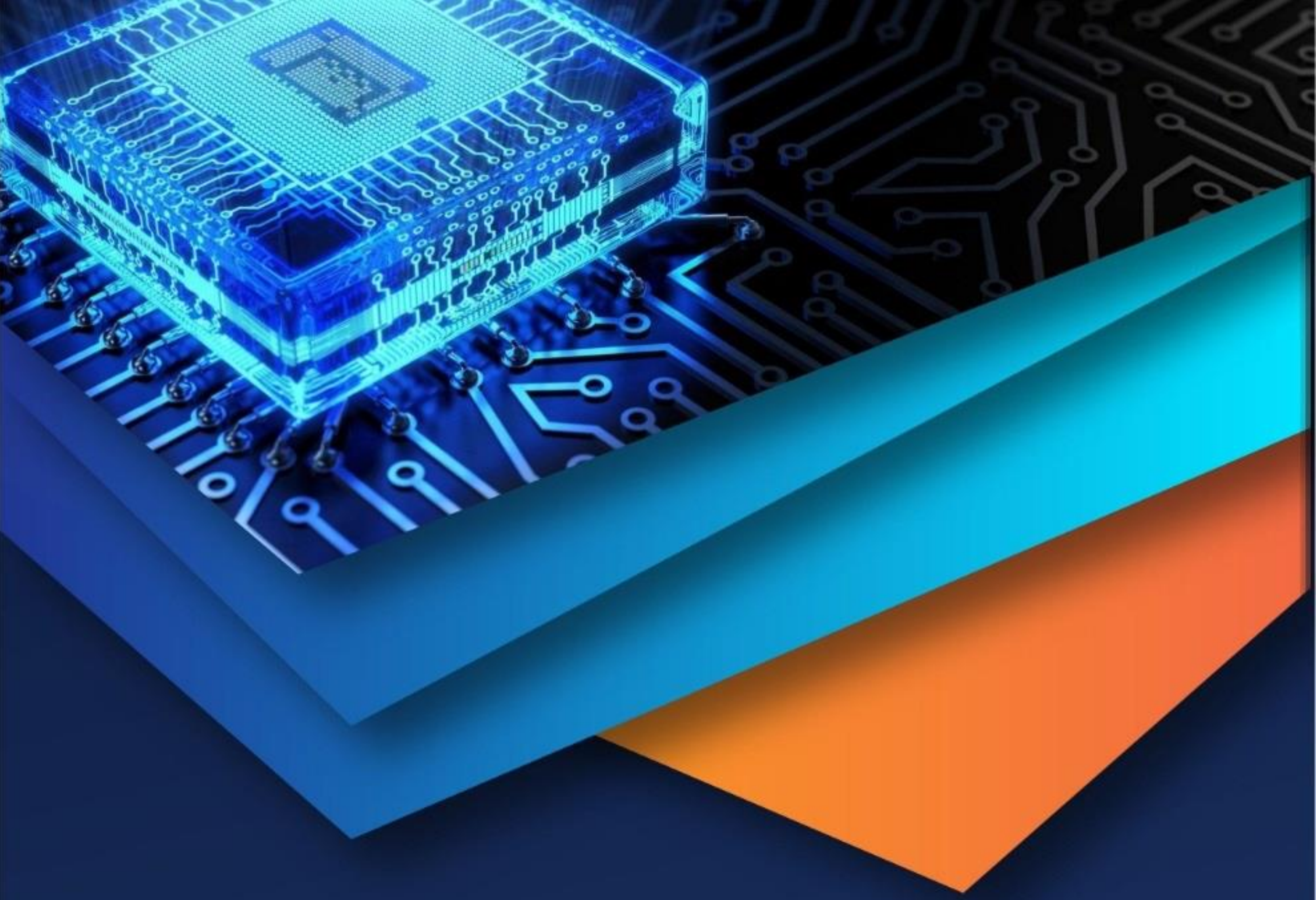
[18] V. Jain, M. Agrawal and A. Kumar, "Performance Analysis of Machine Learning Algorithms in Credit Cards Fraud Detection," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2020, pp. 86-88, doi: 10.1109/ICRITO48877.2020.9197762.

[19] R. Qaddoura and M. M. Biltawi, "Improving Fraud Detection in An Imbalanced Class Distribution Using Different Oversampling Techniques," 2022 International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI), Zarqa, Jordan, 2022, pp. 1-5, doi: 10.1109/EICEEAI56378.2022.10050500.

[20] R. Roy and K. T. George, "Detecting insurance claims fraud using machine learning techniques," 2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT), Kollam, India, 2017, pp. 1-6, doi: 10.1109/ICCPCT.2017.8074258.

[21] A. Shivanna, S. Ray, K. Alshouiliy and D. P. Agrawal, "Detection of Fraudulence in Credit Card Transactions using Machine Learning on Azure ML," 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 2020, pp. 0268-0273, doi: 10.1109/UEMCON51285.2020.9298129.

[22] K. J and A. Senthilselvi, "Credit Card Fraud Detection based on Ensemble Machine Learning Classifiers," 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022, pp. 1604-1610, doi: 10.1109/ICESC54411.2022.9885649.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY