iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59045

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Examining Various ML Approaches in Blockchain for Fraud Detection

T. Nihith Novah[1], M. Varun[2], K. Pavan Reddy[3], K. Ragini[4]
[1, 2, 3]*UG Student, Department of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana*
[4]*Professor, Department of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana*

Abstract: ***Fraudulent transactions significantly impact blockchain network trust and the economy. Traditional consensus methods (e.g., proof of work or proof of stake) can't confirm the identity of participants, leaving the network susceptible to fraud. Machine learning algorithms offer a potential solution to detect fraudulent transactions and participants. Fraudulent exchanges in the blockchain economy deter investors and raise skepticism. This study explores the effectiveness of controlled AI and deep learning models in identifying fraudulent transactions and users, integrating machine learning with blockchain technology.***
Keywords: Blockchain, Digital Currency, Transactions, Decentralized Network, Fraudulent, Proof of work, Peer-to-Peer Transactions.

I. INTRODUCTION

The persistent issue of identifying fraudulent transactions within blockchain networks has garnered substantial attention over time. Such transactions not only pose a threat to the economy but also undermine trust in cryptocurrencies like bitcoins and other blockchain-based solutions. Fraudulent activities typically raise suspicion either due to the parties involved or the transaction characteristics. Members of blockchain networks are keen on swiftly identifying fraudulent transactions to safeguard the community and preserve the network's integrity. Numerous Machine Learning (ML) techniques have been proposed to tackle this challenge, yielding promising results. However, there lacks a definitive superior method among them. This paper endeavors to compare the efficacy of various supervised ML models, including Support Vector Machines (SVM), Decision Trees, Naive Bayes, Logistic Regression, and several deep learning models, in detecting fraudulent transactions within blockchain networks. Such a comparative analysis aims to discern the optimal algorithm by balancing accuracy and computational speed. The overarching objective is to pinpoint users and transactions with the highest likelihood of involvement in fraudulent activities.

II. EXISTING SYSTEM

1) Synthetic data generation using SMOTE oversamples malicious entities, reducing classification bias. Results, depicted for balanced dataset. Observing log loss during XGBoost training. The small gap between training and test data log loss suggests the model's capability for real-world anomaly detection in blockchain networks.
2) Blockchain technology is a powerful tool for preventing fraud in business networks. It creates an unalterable transaction record, ensures data security, and addresses privacy concerns by anonymizing data and enforcing permission-based access. The consensus process adds an extra layer of validation before transactions are added to the blockchain.
3) Preprocessing and Data Handling Model Building and Evaluation Various machine learning and deep learning techniques are employed to predict transaction success. Models are evaluated using bootstrapping to estimate parameters and determine efficacy based on mean accuracy Values.

III. LITERATURE SURVEY

1) In her research conducted in 2016, Xu delved into the vulnerabilities of blockchain technology to malicious attacks, emphasizing the distinction between identifiable fraudulent activities and those that persist as challenges. She underscored the limitations of blockchain in detecting sophisticated attacks like identity theft and system hacking, which exploit its reliance on predefined rules. Xu's findings underscore the need for the integration of machine learning solutions to enhance the security posture of blockchain systems in mitigating such threats.
2) In their 2019 study, Shi et al. employed transaction aggregation techniques to analyze customer behavior preceding transactions in the Bitcoin market. They aimed to detect fake transactions by examining customer behavior, developing a model capable of identifying anomalies in unknown datasets, including those provided by banks with privacy concerns. The model treated all participant attributes equally without prioritization, effectively distinguishing between legal and fake transactions within improper datasets.

3) Ostapowicz and Zbikowski (2019) employed Supervised Machine Learning methods to detect fraudulent accounts within blockchain systems. Their approach aimed to combat threats such as malware and fake emails, commonly used by malicious actors to pilfer funds. Using a dataset containing over 300,000 accounts, they applied Random Forests, Support Vector Machines, and XGBoost classifiers to identify and flag fraudulent activities.

4) Apruzzese et al. (2020) addressed the vulnerability of intrusion detection systems in cybersecurity, particularly when utilizing datasets with highly sensitive training data. They proposed hardening random forest cyber detectors averse to adversarial attacks to improve cyber-attack detection. Despite the use of random forest algorithms for enhanced detection, there remains scope for further improvement in detecting cyber-attacks.

IV. METHODOLOGY

A. Classification of Attacks

Blockchain technology is often heralded for its robust security in financial transactions. However, despite its advancements, it remains susceptible to contemporary cyber-attacks. Even though the integration of multiple security measures, certain proficient cybercriminals try to orchestrating potent attacks against blockchain networks. Attacks such Sybil attacks, double-spending attacks, Denial of service attacks and remain terrifying challenges to blockchain security. As a result, these attacks are disturbing the blockchain network adversely.

B. Dataset

The dataset comprises various fields including indexes, transaction timestamps, unique addresses, minimum and maximum contract values, transaction types, etc., which are utilized to analyze transaction histories and determine whether transactions are legitimate or fraudulent. It encompasses nearly 9746 records. Using these dataset records, we assess the accuracy, precision, F-score, and recall of different algorithms to determine which algorithm performs the best.

C. Data Analysis

Exploratory Data Analysis (EDA) is a fundamental step in understanding and comprehensively analyzing a dataset. It involves several key tasks aimed at gaining insights into the data's structure and characteristics. Initially, assigning meaningful column names is crucial as it provides important identifiers to each attribute, facilitating easier interpretation and analysis. Subsequently, validating for invalid values ensures that there are no erroneous entries within the dataset, which could otherwise distort analysis outcomes or impede display performance. Moreover, data visualization techniques such as creating plots and charts are employed to visually represent the distribution of data and explore relationships between different features. These visualizations aid in identifying patterns, trends, and anomalies within the dataset, thereby informing subsequent steps in the data analysis process. In our project, it's crucial to factor to know the transaction is fraud or legitimate and verify the transaction history.

D. Algorithms

In our extent, we utilize a differing extend of calculations to address different angles of our issue. Here the data is divided into two sets train (0.8) and test (0.2). The ratio of fraudulent to legitimate transactions is then verified in our train and test sets. By using different ML and Deep Learning models to detect whether the transaction is fraudulent or not. They are Logistic Regression, MLP, Naive Bayes, AdaBoost, Decision Tree, SVM, Random Forest, Deep Network.

E. Implementation Block Diagram

Machine learning algorithms have become increasingly popular in various domains due to their ability to extract insights from data and make predictions. However, implementing these algorithms effectively requires careful consideration of several factors, including data preprocessing, model training, algorithm selection, and performance evaluation. This research paper presents a framework that addresses these challenges by providing a structured approach to implementing machine learning algorithms on datasets. The framework consists of four main modules: Upload & Preprocess Dataset, Generate Train & Test Model, Algorithm Execution and Performance Evaluation, and Comparison Graph. Each module plays a crucial role in different stages of the implementation process, ensuring a systematic and efficient approach to building machine learning models. The Upload & Preprocess Dataset module allows users to upload their datasets and perform preprocessing tasks. This module is essential for ensuring data quality and consistency before training the models. It includes functionality to handle missing values, a common issue in real-world datasets, which can significantly impact model performance if not addressed properly.

The Generate Train & Test Model module facilitates dataset splitting into training and testing subsets. It provides users with essential information about the dataset, such as the total number of records and columns, enabling them to gain insights into the data's characteristics. This module sets the foundation for model training and evaluation by preparing the data for further analysis. The Algorithm Execution and Performance Evaluation module allow users to run various machine learning algorithms and assess their performance. Users can execute different algorithms and compare their accuracy scores to determine the most suitable model for their dataset. This module provides valuable insights into the strengths and weaknesses of each algorithm, enabling informed decision-making in algorithm selection. The Comparison Graph module generates visual representations of algorithm performance, facilitating easier analysis and interpretation of results. By visualizing the accuracy scores of different algorithms, users can identify trends and patterns that may not be apparent from numerical data alone. This module enhances the overall usability of the framework by providing intuitive visualizations for better decision-making.

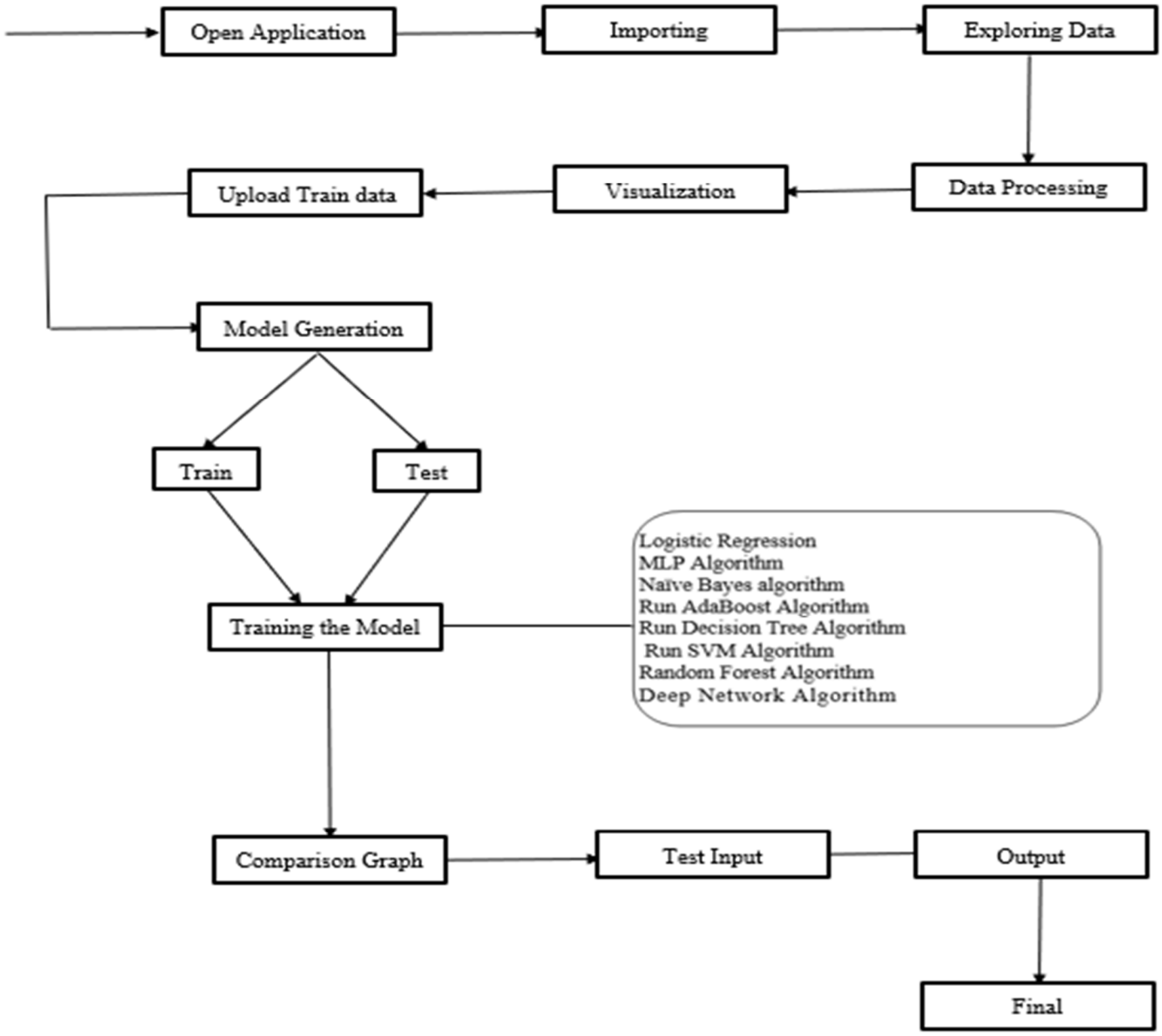
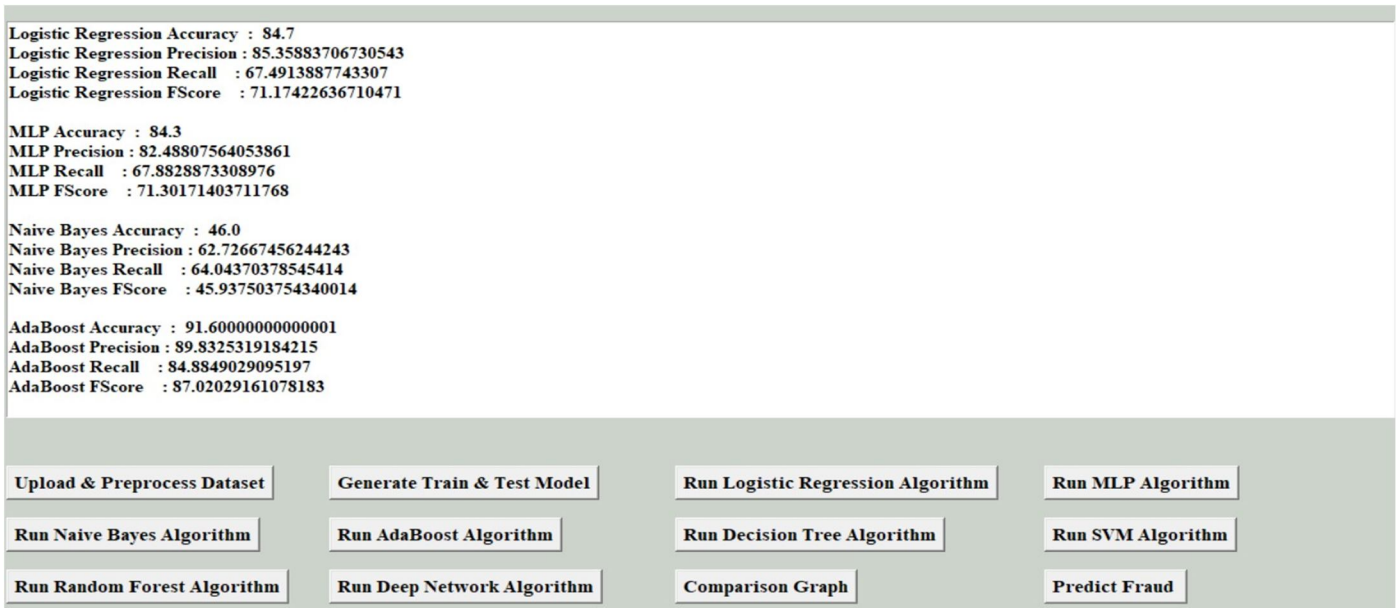


Fig. Architecture of the proposed model

V. PERFORMANCE

A. *Performance Validation Of Machine Learning Algorithms*



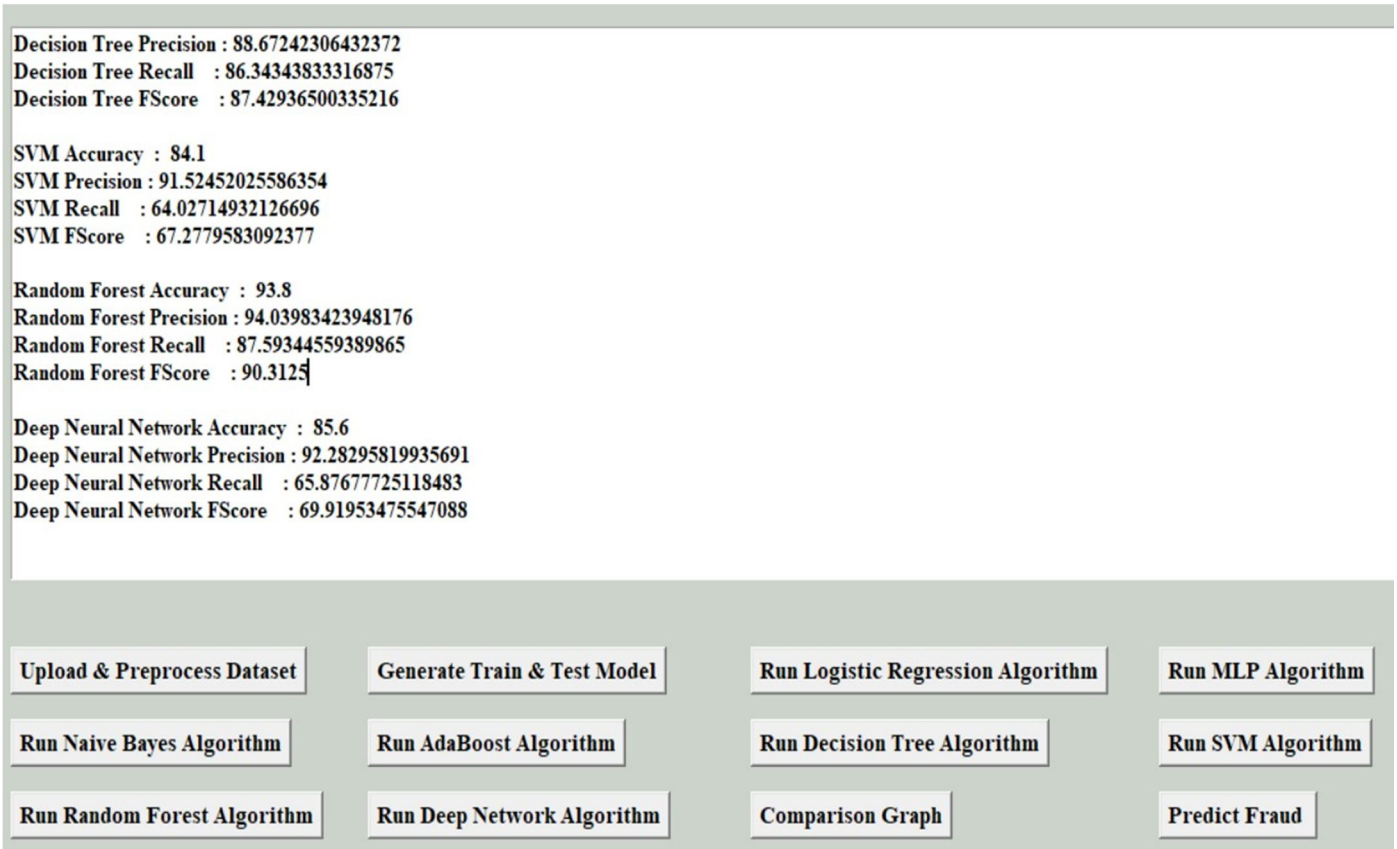


Fig 2: Performance Metrics 2

VI. RESULTS AND DISCUSSION

A. *Comparision Algorithm*

This Table interprets the data of the comparison of the accuracy, precision, recall and fscore of Logistic Regression, MLP, Naïve Bayes, AdaBoost, Decision Tree, SVM, Random Forest and Deep Neural Network.

Algorithm Name	Accuracy	Precision	Recall	FSCORE
Logistic Regression Algorithm	83.2	86.80351551638681	66.54978771520676	69.80828259447492
MLP Algorithm	83.5	83.80955209694665	68.56418285304238	71.86945380708582
Naive Bayes Algorithm	49.6	61.845714912761984	63.497344401498225	49.44144838212634
AdaBoost Algorithm	92.10000000000001	90.04081772655553	88.07590064745503	88.99781209655744
Decision Tree Algorithm	91.9	89.031051701956	88.92177808220667	88.97621720934981
SVM Algorithm	82.5	90.61158798283262	63.991769547325106	66.68437154350354
Random Forest Algorithm	94.19999999999999	94.06772025300488	89.882088164783	91.7365262635991
Deep Neural Network Algorithm	84.8	91.16875634864083	65.77291922002335	69.52295402778223

Table 1: Comparison Table

B. *Figures*

It represents the comparison bar graph of various Machine learning algorithms. X-axis represents the different algorithms like AdaBoost, Decision Tree, MLP, Random Forest, SVM, Naive Bayes, Logistic

We can have a clear view that interms of accuracy and f-score Random Forest has the high performance and in we can have a clear view that interms of accuracy and f-score of Random Forest is best out of all other algorithms.

Regression and Deep Neural Network Algorithm. Y-axis represents the scale of each algorithm. Here, we are calculating Accuracy, Precision, F1 score and recall for each algorithm to prove which is performing best.

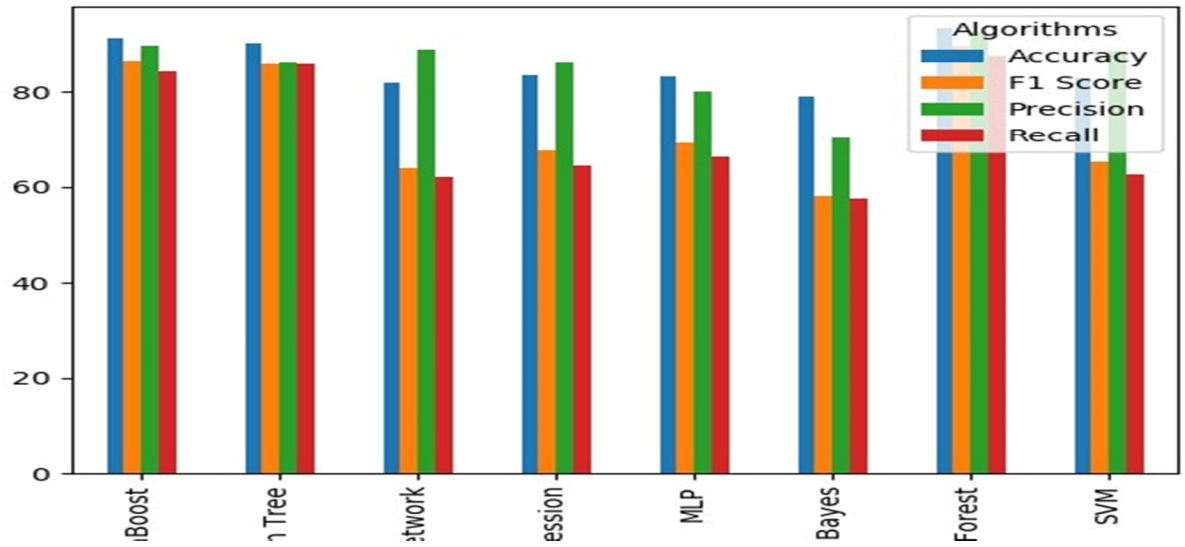


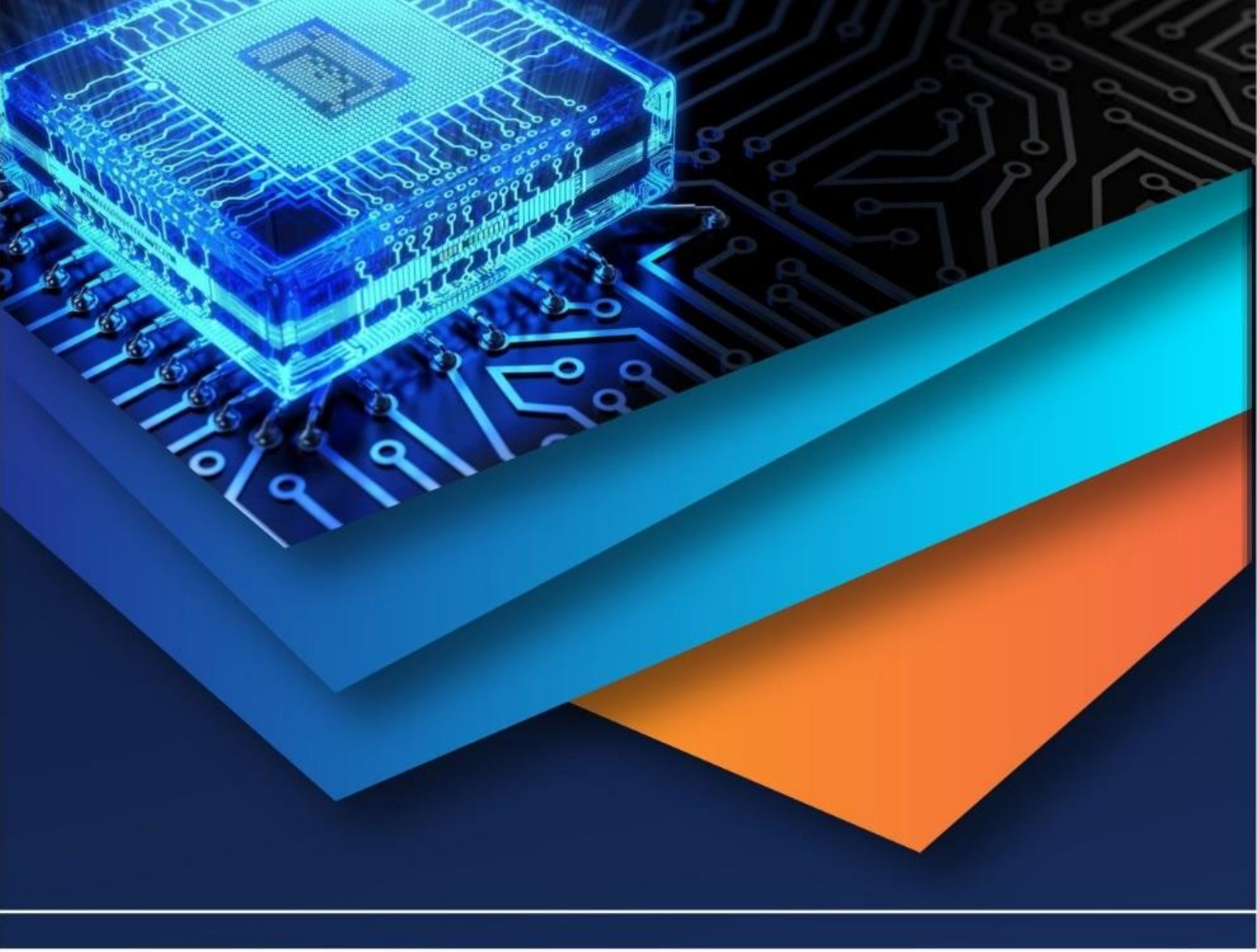
Fig. 3: Graphical representation of Comparison of models

VII. CONCLUSION

A study explored the use of machine learning to detect fraudulent transactions within a blockchain network. The researchers examined various supervised learning algorithms like support vector machines, decision trees, logistic regression, and dense neural networks, comparing their performance in terms of accuracy. They suggested extending the study to include unsupervised techniques such as clustering. Additionally, the researchers expressed intentions to conduct a detailed investigation into fraudulent activities within a private blockchain network in the future.

REFERENCES

[1] Zero-knowledge proof-of-identity: Sybil-resistant, anonymous authentication on permissionless blockchains and incentive compatible, strictly dominant …DC Sánchez - arXiv preprint arXiv:1905.09093, 2019 - arxiv.org
[2] Ensuring consensus on trust issues in capability-limited node networks with Blockchain technology S Hadjiefthymiades, M Chatzidakis, D Reisis - pergamos.lib.uoa.gr
[3] Comparative study on identity management methods using blockchain AG Nabi - University of Zurich, 2017 - files.ifi.uzh.ch
[4] The blockchain and the new architecture of trust K Werbach - 2018 - books.google.com
[5] Effectiveness of Machine and Deep Learning for Blockchain Technology in Fraud Detection and Prevention Y Kumar, S Gupta - Applications of Artificial Intelligence, Big Data …, 2022 - taylorfrancis.com
[6] Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019 KG Al-Hashedi, P Magalingam - Computer Science Review, 2021 – Elsevier
[7] Analyzing Various Machine Learning Algorithms for Blockchain-Based Fraud Detection S Giribabu, V Sriharsha, PH Basha… - … Research in …, 2022 - acspublisher.com
[8] https://learnprompting.org/docs/basics/instructions
[9] https://chat.openai.com/
[10] https://bard.google.com/chat



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY