iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.59899

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 IV Apr 2024- Available at www.ijraset.com

Fraud Detection on Bank Payments

Dr. D. Suneetha[1], M. Bhagavat[2], Gollu Siddu[3], CH. Mahesh[4], D. Khadyoth[5], P. Jahnavi[6]
[1]*Assistant Professor, Department of computer science, GITAM School of Technology, Visakhapatnam-530045, Andhra Pradesh*
[2, 3, 4, 5, 6]*Students, Department of computer science, GITAM School of Technology, Visakhapatnam-530045, Andhra Pradesh*

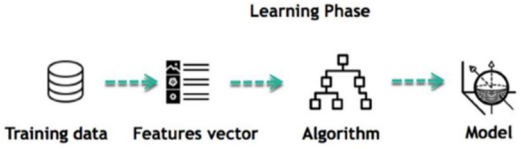
Abstract: ***A "financial fraud" occurs when money is obtained via dishonest and illegal ways. The use of deceitful means to get financial benefits, or financial fraud, has lately grown to be a major concern for organizations and corporations. Despite several efforts to reduce it, financial fraud continues to harm both society and the economy, causing huge losses every day. Antiquity is the cradle of several techniques for detecting deceitful acts.***
Handiwork is the norm, despite its many drawbacks: it's time-consuming, costly, prone to mistakes, and inefficient. No research has been able to decrease fraud-related losses thus far, but there may be more on the way. Traditional approaches to identifying these fraudulent operations rely on labor-intensive, costly, and prone-to-error human verifications and inspections.
Recent developments in artificial intelligence (AI) have made it possible to efficiently examine massive amounts of financial data for indications of fraud using methods based on machine learning. This research fills that information gap by developing a new model for detecting fraudulent bank payments using the Random Forest Classifier Machine Learning Algorithm. Our proposed strategy outperforms the existing one, as shown by a train/test accuracy rate of 99% on the Banksim dataset.

I. INTRODUCTION

Instead than being hard-coded from the beginning, what we refer to as "machine learning" is really a set of algorithms that computers may execute that can learn from their own observations and experiences. Machine learning is a branch of artificial intelligence that makes use of statistical techniques to analyze data and make predictions that might be useful for decision-making. The original idea stems from the possibility that computers may learn from data samples and produce accurate results on their own. There is a tight relationship between machine learning, data mining, and Bayesian predictive modeling. The machine takes in data, processes it using an algorithm, and finally produces a result. One major challenge in machine learning is making recommendations. Individual users' viewing patterns form the basis of all of Netflix's movie and TV program selections. Unsupervised learning is being used by IT organizations to improve the customer experience via tailored recommendations. Machine learning has several uses, including process automation, portfolio optimization, maintenance demand forecast, and fraud detection.

A. How Is Machine Learning Conducted?

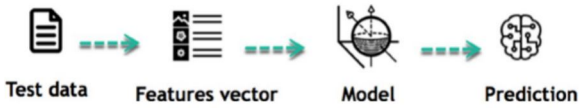
In terms of learning, machine learning is analogous to possessing an extra cerebral cortex. A computer's learning process is quite similar to that of a human brain. Doing things helps people learn. Better forecasts will be possible after we have more data. As an example, our odds of success are lower in a situation where we don't know the outcome than in one where we do. The same thing is taught to all machines. Before the computer can make an accurate prediction, it needs to see a sample. If we give the computer a similar example, it can figure it out. But when faced with an unseen example, computers' prediction abilities are no different from human ones.
Machine learning revolves on learning and inference. Pattern recognition is the primary means by which the computer learns. This discovery is the result of the data. The ability to intelligently choose which data points to send into machines is a crucial function of data scientists. A problem may be addressed by using a feature vector, which is a collection of attributes. The feature vector is analogous to a data subset that is used to address a problem.
After learning this new information, the computer uses complicated algorithms to create a model that makes everything easier to understand. Consequently, the learning process involves describing and summarizing data into a model.



©IJRASET: All Rights are Reserved

For example, the algorithm is trying to figure out whether there's a correlation between a person's income and their tendency to eat at fancy restaurants. A visit to a high-end restaurant is positively associated with income, as shown by the machine: Look at this, the model!

Inferring: After the model is constructed, its efficacy may be evaluated using unique data sets. A features vector is generated from the updated data, and then before making a forecast based on the model. The most appealing aspect of machine learning is this whole thing. Neither the rules nor the model need to be retrained. When you have trained a model, you may use it to draw conclusions from fresh data.



The following are the main aspects of a Machine Learning program's lifecycle:

1) Define a question
2) Collect data
3) Visualize data
4) Train algorithm
5) Collect feedback
6) Refine the algorithm
7) Loop 4-7 until the results are accurate
8) Use the model to make prediction

The system learns from its mistakes and applies its findings to fresh data sets.

II. LITERATURE SURVEY

A. Building a robust mobile payment fraud detection system with adversarial examples

AUTHORS:S. Delecourt and L. GuoMobile payment is becoming a major payment method in many countries. However, the rate of payment fraud with mobile is higher than with credit card. One potential reason is that mobile data is easier to be modified than credit card data by fraudsters, which degrades our data-driven fraud detection system. Supervised learning methods are pervasively used in fraud detection. However, these supervised learning methods used in fraud detection have traditionally been developed following the assumption that the environment is benign; there are no adversaries trying to evade fraud detection systems. In this paper, we took potential reactions of fraudsters into consideration to build a robust mobile fraud detection system using adversarial examples. Experimental results showed that the performance of our proposed method was improved in both benign and adversarial environments.

B. Importance of smart meters data processing – case of Saudi Arabia

AUTHORS: T. Alquthami, A. M. Alsubaie, and M. Anwer

This paper presents a thorough analysis of 30-minute data sets of KSA residential digital meters to identify all possible discrepancies in the data sets and devise statistical techniques best suited to remove these discrepancies as per the nature of each discrepancy. The analysis is performed through a program that was developed in Python-Pandas. The program parses through three month's meter measurements of 3,283 consumers throughout KSA and detects data inconsistencies, duplicates, missing and outlier values and other issues in the data sets. Statistical techniques that are part of the program are then implemented to correct for these issues. A validation process was developed and included in the program to ensure the adjustment process produces the best reliable outcomes. Analysis indicates that smart meters data have issues that need preprocessing to be used for other applications. The outcome of the program developed shows that smart meters measurement outcome data set could be considered as valid and trusted, which can be used for smart grid applications such as behavioral analysis of the electricity consumers.

C. Comparative evaluation of credit card fraud detection using machine learning techniques

AUTHORS: O. Adepoju, J. Wosowei, S. lawte, and H. Jaiman Credit card fraud is a serious and growing problem with the increase in e-commerce and online transactions in this modern era. With this identity theft and loss of money, such mischievous practices can affect millions of people around the world.

Criminal activity is a rising threat to the financial sector with-reaching implications. Information extraction seemed to have assumed a basic job in recognition of online payment fraud, fraud detection efficiency in credit card purchases is significantly affected by the data set measuring strategy, the choice of variable and the detection techniques used. This publication inspects execution of, Support Vector Machine, Naive Bayes, Logistic Regression and K-Nearest Neighbor on exceptionally distorted data on credit card fraud. The execution of these techniques is assessed dependent on accuracy, sensitivity, precision, specificity. The outcomes show an ideal accuracy for logistic regression, Naive Bayes, k-nearest neighbor and Support vector machine classifiers are 99.07%, 95.98%, 96.91%, and 97.53% respectively. The relative outcomes demonstrate that logistic regression performs superior to other algorithms.

D. Supervised machine learning algorithms for credit card fraud detection: A comparision

AUTHORS: S. Khatri, A. Arora, and A. P. AgrawalIn today's economic scenario, credit card use has become extremely commonplace. These cards allow the user to make payments of large sums of money without the need to carry large sums of cash. They have revolutionized the way of making cashless payments and made making any sort of payment convenient for the buyer. This electronic form of payment is extremely useful but comes with its own set of risks. With the increasing number of users, credit card frauds are also increasing at a similar pace. The credit card information of a particular individual can be collected illegally and can be used for fraudulent transactions. Some Machine Learning Algorithms can be applied to collect data to tackle this problem.This paper presents a comparison of some established supervised learning algorithms to differentiate between genuine and fraudulent transactions.

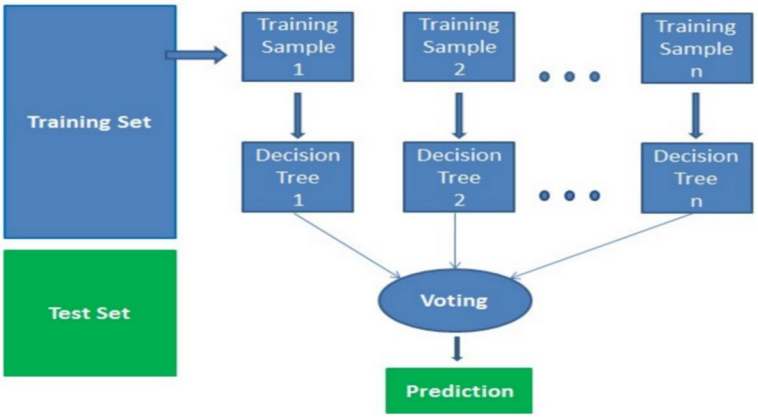
E. Performance analysis of machine learning algorithms in credit cards fraud detection

AUTHORS: V. Jain, M. Agrawal, and A. Kumar Credit cards are very commonly used in making online payments. In recent years' frauds are reported which are accomplished using credit cards. It is very difficult to detect and prevent fraud which is accomplished using credit cards. Machine Learning(ML) is an Artificial Intelligence (AI) technique which is used to solve many problems in science and engineering. In this paper, machine learning algorithms are applied on a data set of credit card frauds and the power of three machine learning algorithms is compared to detect the frauds accomplished using credit cards. The accuracy of Random Forest machine learning algorithm is best as compared to Decision Tree and XGBOOST algorithms.

III. PROPOSED SYSTEM

It is becoming more difficult for banks to detect fraudulent bank payments. Machine learning is vital for detecting fraudulent financial transactions. The proposed approach makes use of machine learning techniques to foretell these transactions by analyzing past data and improving prediction power by adding new features. The approach utilized for data sampling, variable selection, and detection processes has a substantial influence on the efficacy of fraud detection in financial transactions. Kaggle provided us with the dataset of monetary transactions.

The proposed technique successfully identifies fraudulent transactions in the Banksim dataset. The information in question is artificially created and contains payments from several customers across a wide range of time periods and amounts. Applying a Random Forest Classifier model to the dataset follows the data preparation. We evaluate the models' efficacy by observing their performance after training and testing. We compare and analyze the accuracies to get the best accurate model. Our proposed model for the system had a 99% success rate throughout testing and training.



A. *Advantages Of Proposed System*

1) The proposed system reduces overfitting in decision trees and helps to improve the accuracy.
2) It is flexible to both classification and regression problems.
3) It works well with both categorical and continuous values.
4) It automates missing values present in the data.
5) Normalizing of data is not required as it uses a rule-based approach.
6) Decreased the total number of verification steps and measures
7) Real-time processing
8) The proposed system produced good predictions that can be understood easily.
9) It can handle large datasets efficiently.
10) The proposed system provided a higher level of accuracy in predicting outcomes over the decision tree algorithm.

IV. IMPLEMENTATION

1) *Step-1: Data collection*

The dataset used for this project, "Fraud Detection on Bank Payments," is collected from a local directory path specified in the code. Specifically, the dataset file named "fraud.pkl" is loaded using Python's pickle.load() function. The dataset presumably contains features related to bank payments, which are utilized for training a machine learning model to detect fraudulent transactions. It's important to ensure that the dataset file path is correctly specified in the code.

```
import pickle

fraud = pickle.load(open('/Users/pineapple/Desktop/PROJECT PHASE/Fraud Detection /
```

2) *Step-2: Data preprocessing*

Upon loading the dataset, it is expected that the features and labels are appropriately structured for model training. However, specific preprocessing steps may be applied based on the nature of the dataset. For instance, missing value imputation, feature scaling, or encoding categorical variables might be necessary. In this code snippet, the preprocessing steps are not explicitly mentioned, but they may have been performed earlier during dataset creation or exploration.

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC

# Pre-trained model loading
fraud = pickle.load(open('/Users/pineapple/Desktop/PROJECT PHASE/Fraud Detection /
```

3) *Step-3: Model initialization and training*

The provided code snippet focuses on building a Flask web application for fraud prediction using a pre-trained machine learning model. Therefore, the model training process isn't explicitly outlined here. Instead, a pre-trained model stored in the "fraud.pkl" file is loaded using pickle.load(). This model is assumed to have been trained on a separate dataset using algorithms like DecisionTreeClassifier or SVM (Support Vector Machine) for fraud detection based on bank payment features.

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC

# Pre-trained model loading
fraud = pickle.load(open('/Users/pineapple/Desktop/PROJECT PHASE/Fraud Detection
```

4) Step-4: Web application setup and prediction

The Flask web application is initialized, and various routes are defined for different functionalities like rendering HTML templates, uploading datasets, displaying previews, and making predictions. When a user uploads a dataset through the web interface, the uploaded dataset is read and displayed for preview. Subsequently, when the user submits data for prediction, the features are extracted from the form, and the pre-trained model predicts whether the transaction is fraudulent or benign. The prediction result is then displayed to the user on the web interface.

```
@app.route('/preview',methods=["POST"])
def preview():
    if request.method == 'POST':
        dataset = request.files['datasetfile']
        df = pd.read_csv(dataset, encoding='unicode_escape')
        df.set_index('Id', inplace=True)
        return render_template("preview.html", df_view=df)

@app.route('/predict', methods=['POST'])
def predict():
    int_feature = [x for x in request.form.values()]
    final_features = [np.array(int_feature)]
    result = fraud.predict(final_features)
    if result == 1:
        result = "Transaction Fraudulent"
    else:
        result = 'Benign'
    return render_template('prediction.html', prediction_text=result)
```

5) Step-5: Further functionalities

The provided code also includes routes for displaying login pages, performance metrics, and charts. However, the implementation details for these functionalities are not provided in the code snippet. Presumably, these routes would render HTML templates containing login forms, performance metrics visualization, and charts related to fraud detection. These functionalities are related to user interface components and possibly additional features for data analysis or visualization. Let's explain each of these functionalities:

A. Performance Metrics Display (/Performance Route)

This route is intended to display performance metrics related to the fraud detection model. It renders an HTML template named "performance. html". In the HTML template, various performance metrics such as accuracy, precision, recall, F1-score, sensitivity, and specificity may be displayed. These metrics are likely calculated based on the model's predictions on a test dataset or through cross-validation during the model training phase.

B. Chart Visualization (/chart route)

This route is designed to render a chart visualization related to fraud detection or model performance. It serves an HTML template named "chart.html", where the chart component is embedded. The chart could represent various aspects of the data or model, such as distribution of fraudulent vs. benign transactions, trend analysis over time, or comparison of different machine learning algorithms' performance. Visualization is an essential aspect of data analysis, helping users to gain insights and understand patterns more intuitively.

```
@app.route('/performance')
def performance():
    return render_template('performance.html')

@app.route('/chart')
def chart():
    return render_template('chart.html')
```

V. RESULT AND ANALYSIS

The code aims to provide a user-friendly interface for fraud prediction based on bank payment data using a pre-trained machine learning model.



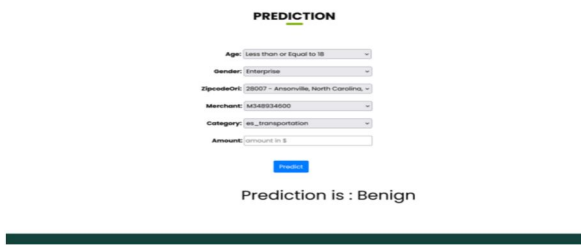












Confusion Matrix: A confusion matrix is a performance measurement technique used in classification tasks to assess the performance of a machine learning model.

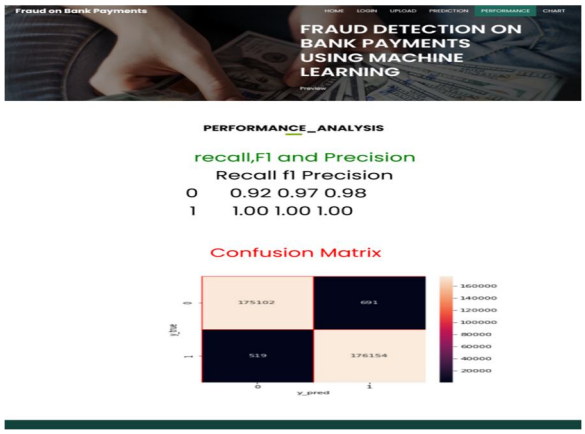
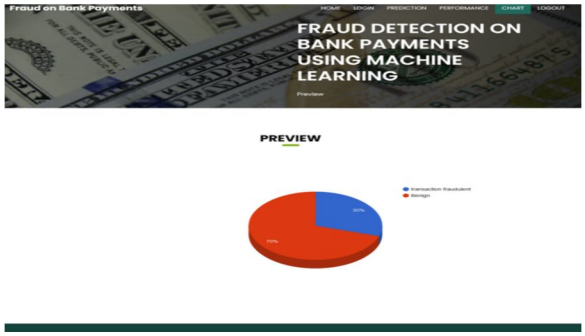


Chart Representation: The purpose of this chart representation is not specified in the code, but it's presumed to provide additional insights into fraud detection or model performance.



VI. CONCLUSION

1) The primary objective of the project is to detect fraudulent transactions in bank payments using machine learning techniques.

2) Through the implementation of the Flask web application, users can upload transaction data, which is then processed and analyzed using a pre-trained machine learning model.

3) The application effectively predicts whether a transaction is fraudulent or benign based on the input features provided by the user.

REFERENCES

[1] S. Delecourt and L. Guo, “Building a robust mobile payment fraud detection system with adversarial examples,” in 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), pp. 103–106, IEEE, 2019.

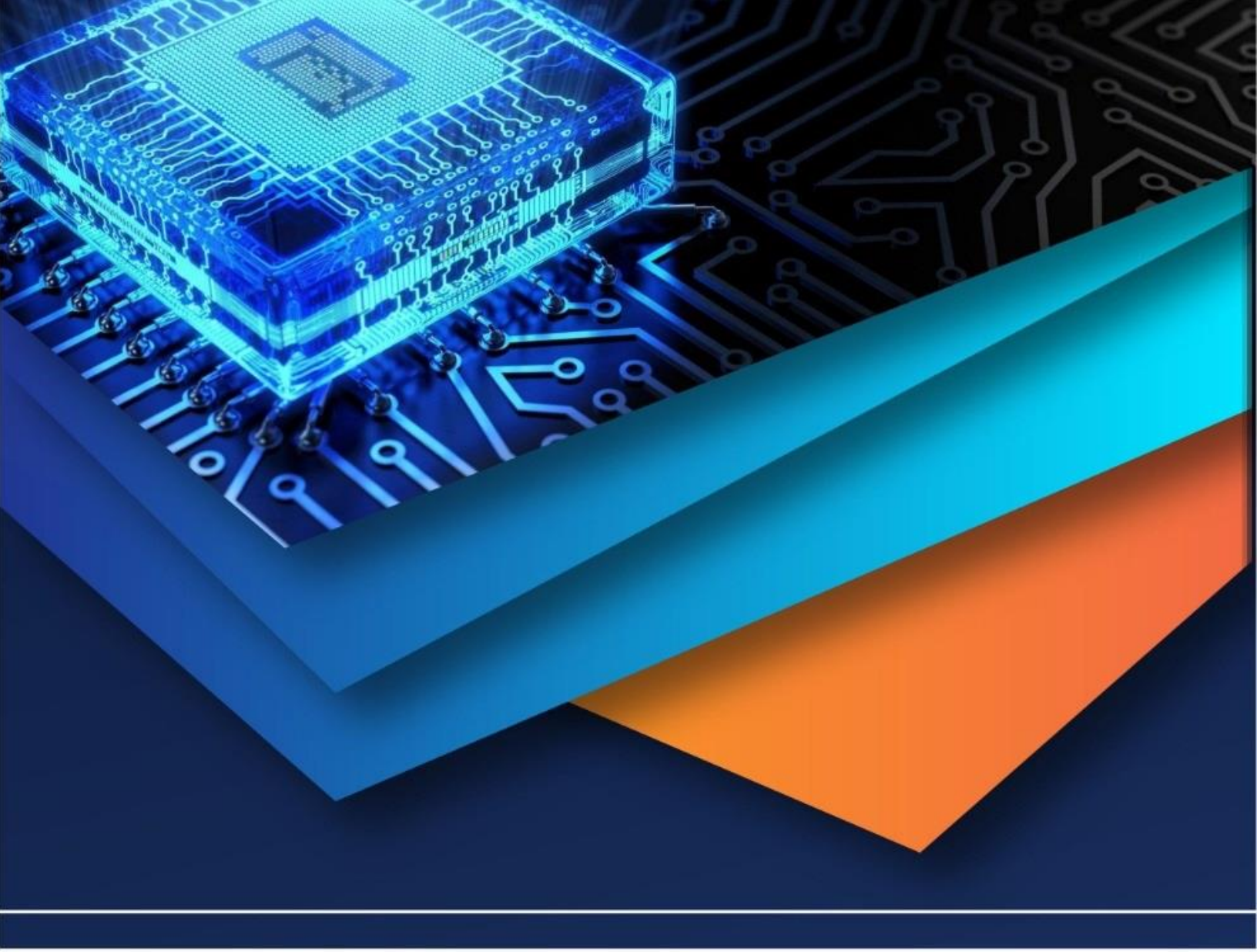
[2] T. Al Qatami, A. M. Alsubaie, and M. Anwer, “Importance of smart meters data processing – case of saudi arabia,” in 2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA), pp. 1–5, IEEE, 2019.

[3] O. Adepoju, J. Wosowei, S. lawte, and H. Jaiman, “Comparative evaluation of credit card fraud detection using machine learning techniques,” in 2019 Global Conference for Advancement in Technology (GCAT), pp. 1–6, IEEE, 2019.

[4] S. Khatri, A. Arora, and A. P. Agrawal, “Supervised machine learning algorithms for credit card fraud detection: A comparison,” in 2020 10th International Conference on Cloud Computing, Data Science Engineering (Confluence), pp. 680–683, IEEE, 2020.

[5] V. Jain, M. Agrawal, and A. Kumar, “Performance analysis of machine learning algorithms in credit cards fraud detection,” in 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 86–88, IEEE, 2020.

[6] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga, and N.Kuruwitaarachchi, “Real-time credit card fraud detection using machine learning,” in 2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence), pp. 488–493, IEEE, 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)