iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IX **Month of publication:** September 2023

DOI: https://doi.org/10.22214/ijraset.2023.55698

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 11 Issue IX Sep 2023- Available at www.ijraset.com

Enhancing Credit Card Fraud Detection Using P-XGBoost: A Comparative Study Classical Machine Learning Techniques

Hera Gulam Waris Ansari[1], Dr. Manoj Dnyaneshwar Patil[2]
[1]*Department of Computer Engineering, Alamuri Ratnamala Institute Of Engginering And Technology, Shahapur*
[2]*Department of Computer Engginering, Rajiv Gandhi Institute Of Technology, Mumbai*

Abstract: ***Ability of debit card companies to detect and prevent fraudulent transactions is crucial to safeguard customers from unauthorized charges. Data Science, particularly Machine Learning, plays a pivotal role in addressing this challenge. This project aims to demonstrate the application of machine learning in Credit debit Fraud Detection by modeling a dataset of past credit card transactions, distinguishing fraudulent ones from legitimate ones. The objective is to achieve a fraudulent transactions while minimizing false positives. Debit Card Fraud Detection is a classic classification issue.***
A research focuses on data analysis, preprocessing, and the utilization of XGBoost on Credit Card Transaction data. To prevent overfitting, grid search is employed to fine-tune the models' hyperparameters. The performed of XGBoost and P-XGBoost is compared with further usual machine learning techniques. Surprisingly, P-XGBoost best XGBoost in fraud detection, presenting a viewpoint for effectively identifying fraudulent behavior while ensuring the privacy of clients.
Keywords: Fraud detection, Credit card, XGBoost, P-XGBoost

I. INTRODUCTION

Machine Learning is a subset of AI that empowers computer learn the systems and improve from experience without explicit programming. Its primary objective is to enable computers to learn autonomously from data and make decisions or predictions based on that learning.

Machine learning is a vital elements of the increasing tract of data science. It involves the use of statistical methods and algorithms to train models that can classify data, make predictions, and uncover valuable insights in data excavation projects. This understanding then operate ruling processes in App and profession, potentially influencing metrics key growth. As the volume of big data continues to expand, the demand for data scientists proficient in machine learning is expected to increase. These professionals will be crucial in identifying relevant business questions and the data needed to answer them effectively.

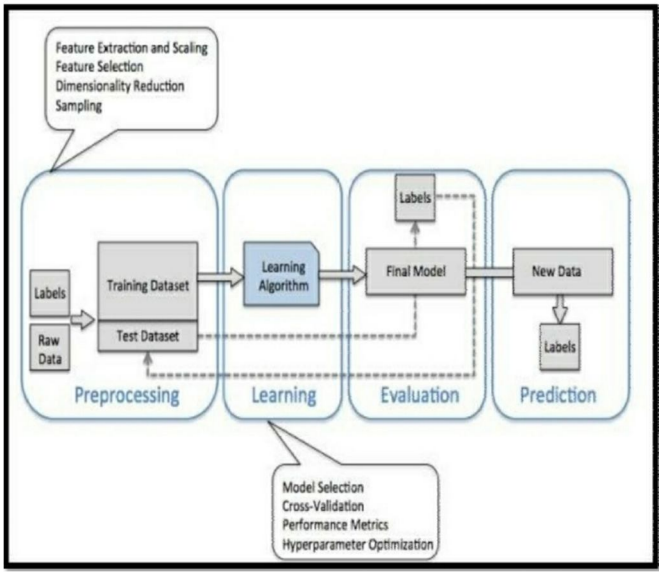


Figure1. Machine Learning

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IX Sep 2023- Available at www.ijraset.com

A. Machine Learning Of Classified

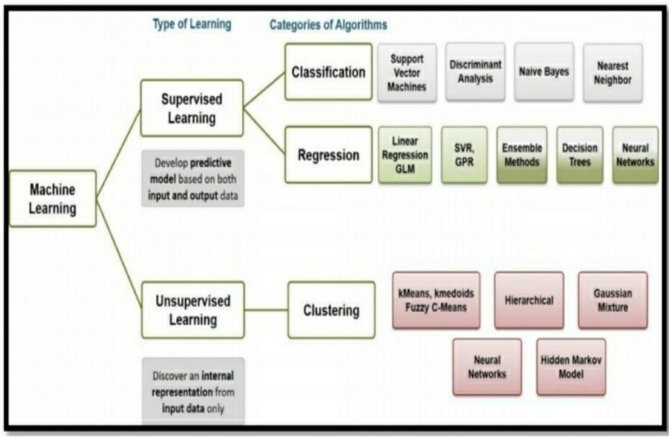


Figure 2. Machine Learning Of Classified

There are several methods of machine learning, classified into specific categories:

1) *Supervised Learning:* Input and output data are provided in this method to the computer during training, along with feedback to evaluate the accuracy of predictions. The goal is to teach the computer how to map input data to the correct output.
2) *nsupervised Learning:* In this approach, no explicit training data with labeled outputs is provided. The computer is left to find patterns or relationships within the data on its own. Unsupervised learning is commonly used in tasks involving transactional data and can be applied to more complex tasks using techniques like deep learning.
3) *Reinforcement Learning:* This type of learning involves three components: the agent, the environment, and actions. The agent perceives its surroundings, interacts with the environment, and takes actions accordingly. The main objective of reinforcement learning is to find the optimal policy for the agent to achieve its goals.

Machine learning has wide-ranging App in various fields, also NLP, machine vision system, admonition systems, fraud detection, and autonomous vehicles, among others. By leveraging the power of data and automated learning, machine learning continues to revolutionize industries and advance technology in exciting ways.

Credit card fraud is a broad term encompassing various theft and fraudulent activities involving the unauthorized use of credit cards during payment transactions. The motivations behind credit card fraud may range from making purchases without payment to transferring unauthorized funds from an account. Credit card fraud often intertwines with identity theft, where thieves use stolen information for fraudulent purposes. According to the United States Federal Trade Commission, identity theft rates remained stable during the mid-2000s but increased by 21 percent in 2008. Despite this increase, credit card fraud, which is commonly associated with identity theft, decreased as a percentage of all identity theft complaints. However, even with improved fraud detection systems, a small percentage of fraudulent transactions still results in significant financial losses for businesses.

Credit card fraudsters employ various methods to execute their fraudulent activities. One common approach is application fraud, where individuals provide false information to obtain a credit card. Another prevalent method involves the unauthorized use of lost or stolen cards. More sophisticated fraudsters may produce fake or altered cards, use skimming techniques to copy card data, or create cloned websites to deceive people into providing their credit card details unknowingly.

Credit card fraud can be classified into different categories, such as online and offline credit card fraud, card larceny, account broke, appliance incursion, App cheating, fake cards, and transport cheating.

The prevalence to credit card fraud highlights the need for robust fraud detection and prevention mechanisms to protect both individuals and businesses from financial losses and security breaches. Effective strategies and technologies are essential to combat this ongoing threat to financial security in today's digital age.

II. LITERATURE SURVEY

1) Clifton Phua and associates conducted a comprehensive survey on fraud detection techniques, including data mining applications, automatic fraud detection, and combatant perception.
2) Suman presented techniques like Supervised and Unsupervised Learning for credit card fraud detection but highlighted the lack of permanent and consistent solutions to fraud detection.

3) Wen-Fang YU and Na Wang used Oddity excavation, oddity perception excavation, or space sum algorithms to correctly forecast fraudulent transactions in an simulation investigation of debot card transaction data.
4) Unconventional method, such as cross data excavation/composite system classifying algorithms, have shown efficiency in perceiving criminal case in card deal data sets, addressing the challenge of strong class imbalance.
5) Artificial Genetic Algorithm was introduced as a new approach to countering fraud, accurately detecting fraudulent transactions while minimizing false alerts, though it faced challenges with varying mistype prices.
6) Various Supervised learning and Semi-Supervised machine learning method have been using for fraud perception, aiming to overcome challenges related to strong class imbalance, labeled and unlabeled samples, and processing a more no of transactions.
7) Different Supervised machine learning algorithms, such as Decision Trees, Naive Bayes Classification, Least Squares Regression, Logistic Regression, and SVM, have been employed for real-time fraudulent transaction detection in datasets.

The literature survey provides an overview of the existing research and techniques employed in the field of credit card fraud detection. Different approaches, algorithms, and challenges associated with fraud detection have been studied, showcasing the ongoing efforts to combat credit card fraud effectively.

III. PROBLEM DEFINITION

Credit card fraud poses a significant threat to businesses, resulting in billions of dollars in losses, despite the implementation of fraud detection systems. The primary challenge in combatting credit card fraud lies in understanding the various methods employed by fraudsters. Credit card fraud occurs when an individual uses someone else's credit card without the owner's or card issuer's knowledge, either by stealing the physical card or obtaining important account information.

The problem at hand involves detecting and preventing fraudulent debit card transactions. Unofficial and undesired usage of an account by someone other than the holder constitutes fraud. The objective is to study the behavior of fraudulent practices and implement prevention measures to minimize such abuses and protect against future occurrences.

Fraud detection entails monitoring user activities to identify objectionable behavior, including fraud, intrusion, and not pay. This is a applicable issue that claim attention from the machine learning and data science communities, as automation can offer potential solutions. However, the problem is challenging due to factors like class imbalance, where valid transactions significantly outnumber fraudulent ones. Transaction patterns also change over time, presenting additional complexities.

Real-world fraud detection systems face a continuous current of amount demands, necessitating rapid scanning by automated compent to allow transactions. Machine learning algorithms are client to analyze allow deal and identify incredulous ones, which are then inquiry into by line of work to confirm their authenticity. The feedback provided by investigators is used to guide and modernize the algorithm, gradually improving cheating perception perform more time.

The ongoing development of fraud detection methods is essential to stay ahead of criminals who continuously adapt their fraudulent strategies. The goal is to create effective and efficient fraud detection systems that protect businesses and consumers from financial losses and maintain the security of credit card transactions.

IV. OBJECTIVE

The primary objective of this research is to implement a robust debit card cheating perception model that can accurately identify fraudulent transactions while minimizing false positives. The problem of debit Card cheating perception is a classic classifying task.

The research procedure focuses on the following key steps:

1) *Data Analysis and Pre-processing:* The first step involves analyzing and pre-processing the credit card transaction datasets. This includes handling missing data, outlier detection, and feature engineering to make sure that the data is satisfactory for trial machine learning models.
2) *Deployment of XG Boost:* XG Boost, a powerful gradient boosting algorithm, is deployed on the pre-processed Credit Card Transaction data. XG Boost is known for its efficiency and effectiveness in handling classification tasks, making it a suitable candidate for this fraud detection problem.
3) *Grid Search for Avoiding Overfitting:* To prevent overfitting, grid search is utilized to tune the hyperparameters of the XG Boost model. This process involves systematically exploring various combinations of hyperparameters to find the optimal settings that maximize the model's performance.

4) *Performance Analysis:* The performance of the XG Boost model is evaluated and compared with other classical machine learning methods. These methods may include Decision Trees, Naive Bayes, SVM, etc., commonly used for fraud detection.
5) *Comparison with P-XGBoost:* Additionally, the research aims to compare the performance of XG Boost with P-XGBoost, an improved version of XG Boost that demonstrates promising results in fraud detection while preserving client privacy.

By achieving the objective of developing an accurate and efficient credit card fraud detection model, this research contributes to enhancing security in financial transactions and mitigating potential financial losses due to fraudulent activities.

V. METHODOLOGY

1) Concept Drift Handling: The methodology aims to address the challenge of concept drift in credit card transactions. Concept drift refers to the changes in transaction patterns over time, leading to unfamiliar and imbalanced data. Techniques will be employed to handle concept drift and adapt the fraud detection model to evolving transaction behaviors.
2) *Data Preprocessing:* The dataset containing credit card transactions made by a cardholder in a two-day period in September 2013 is analyzed. The dataset comprises 284,807 transactions, of which only 492 (0.172percentage) are misleading. The collection of data is extremely disbalanced, with a significantly lower number of fraudulent transactions. To protect customer confidentiality, most property in the collection of data are transformed using Principal Component Analysis (PCA), while 'time', 'payment', and 'class' are non-PCA applied features.
3) *Model Deployment:* The preprocessed dataset is used to train and deploy the fraud detection model. Different machine learning techniques, including XG Boost and P-XG Boost, will be employed to develop the model. These algorithms are well-suited for handling classification tasks and are expected to yield accurate results in detecting fraudulent transactions.
4) *Evaluation Metrics:* Various evaluation metrics such as precision, recall, F1-score, and area under the Receiver Operating Characteristic (ROC) curve will be used to assess the performance of the fraud detection model. These metrics help in measuring the model's accuracy, efficiency, and ability to distinguish between genuine and fraudulent transactions.
5) *Concept Drift Detection:* Techniques for detecting concept drift will be applied during the model's deployment phase. This involves monitoring the changes in transaction patterns over time and updating the model accordingly to adapt to the evolving behavior of transactions.
6) *Comparison and Analysis:* The performance of the developed fraud detection model using XG Boost and P-XG Boost will be compared with other classical machine learning techniques, such as Decision Trees, Naive Bayes, and SVM. The goal is to identify the most effective method for handling concept drift and achieving the highest accuracy in detecting fraudulent transactions.

By combining techniques to handle concept drift, employing advanced machine learning algorithms, and evaluating model performance with appropriate metrics, the methodology aims to create a robust and efficient credit card fraud detection system capable of addressing real-world challenges in financial transactions.

Table 1: Raw features of credit card transactions

Attribute name	Description
Transaction id	Identification number of a transaction
Cardholder id	Unique Identification number given to the cardholder
Amount	Amount transferred or credited in a particular transaction by the customer
Time	Details like time and date, to identify when the transaction was made
Label	To specify whether the transaction is genuine or fraudulent

Figure3. Raw Features Of Credit Card Transactions

Table 2: Attributes of European dataset

S. No.	Feature	Description
1.	Time	Time in seconds to specify the elapses between the current transaction and first transaction.
2.	Amount	Transaction amount
3.	Class	0 - not fraud 1 – fraud

Figure4. Attributes Of European Dataset

The prototype model is a software development approach where a throwaway prototype is built to gain a better understanding of the system requirements before proceeding with the final design and coding. The basic idea is to involve the client in the development process early on and allow them to interact with the prototype to get an "actual feel" of the system.

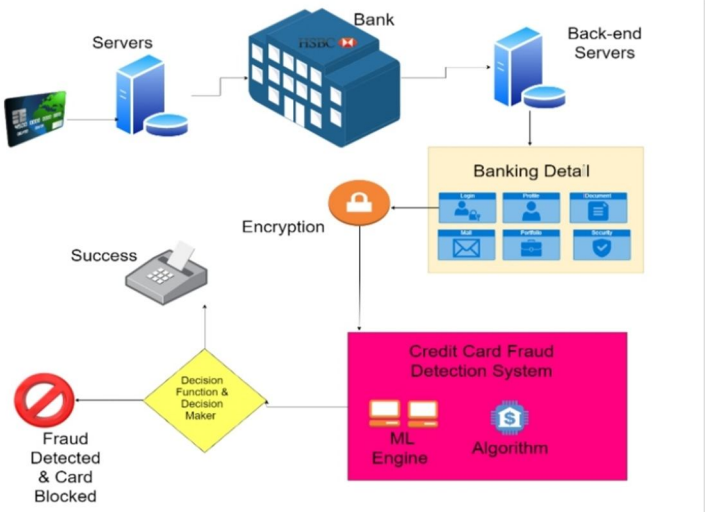


Figure5. System Architecture

A. Key Points of the Prototype Model

1) *Early Feedback:* The prototype is developed based on the currently known requirements, allowing stakeholders to provide early feedback and make changes to the requirements if necessary.
2) *Understanding Requirements:* The prototype helps in clarifying and refining the system requirements as the client can visually see and interact with the prototype, leading to a better understanding of what the final system should entail.
3) *Suitable for Complex Systems:* Prototyping is particularly useful for large and complex systems where requirements are not well-defined or there is no existing system to guide the development process.
4) *Not Complete Systems:* Prototypes are usually not fully functional systems and may lack some details. The goal is to provide an overview of the system's overall functionality rather than implementing all features in the initial prototype.
5) *Iterative Approach:* The development process in the prototype model is iterative. Based on the feedback received, the prototype can be refined, updated, or even discarded if necessary. This iterative process continues until the final system meets the desired requirements.
6) *Client Collaboration:* The prototype model encourages active client involvement throughout the development process, fostering better communication and understanding between developers and clients.

While the prototype model offers benefits like early feedback and enhanced requirement understanding, it also has some limitations. For instance, the focus on prototyping may lead to scope creep or delays in finalizing the system requirements. Additionally, the cost and effort involved in building and discarding prototypes must be carefully managed.

Overall, the prototype model is a valuable approach in situations where requirements are uncertain or complex, as it enables stakeholders to visualize and refine the system's functionality before proceeding with full-scale development.

VI. TECHNOLOGY

1) *Python:* Python is the main programming language used in the project. It is used for various purposes, including web scraping to collect data, processing and analyzing the data, and developing the web scraper. Python's syntax is known for being concise and expressive, allowing programmers to express concepts in fewer lines of code. It is widely used in data science and web development due to its versatility and extensive libraries like 'urllib2' that facilitate web scraping.
2) *MS Excel:* Microsoft Excel is a spreadsheet application used to visualize and process data in the project. Excel provides features for calculations, graphing, pivot tables, and has a macro-programming language (VBA) for custom functions and automation. Excel's user-friendly interface makes it popular for data manipulation and analysis tasks. It is used to clean and format the collected data and create visualizations for better understanding and presentation.

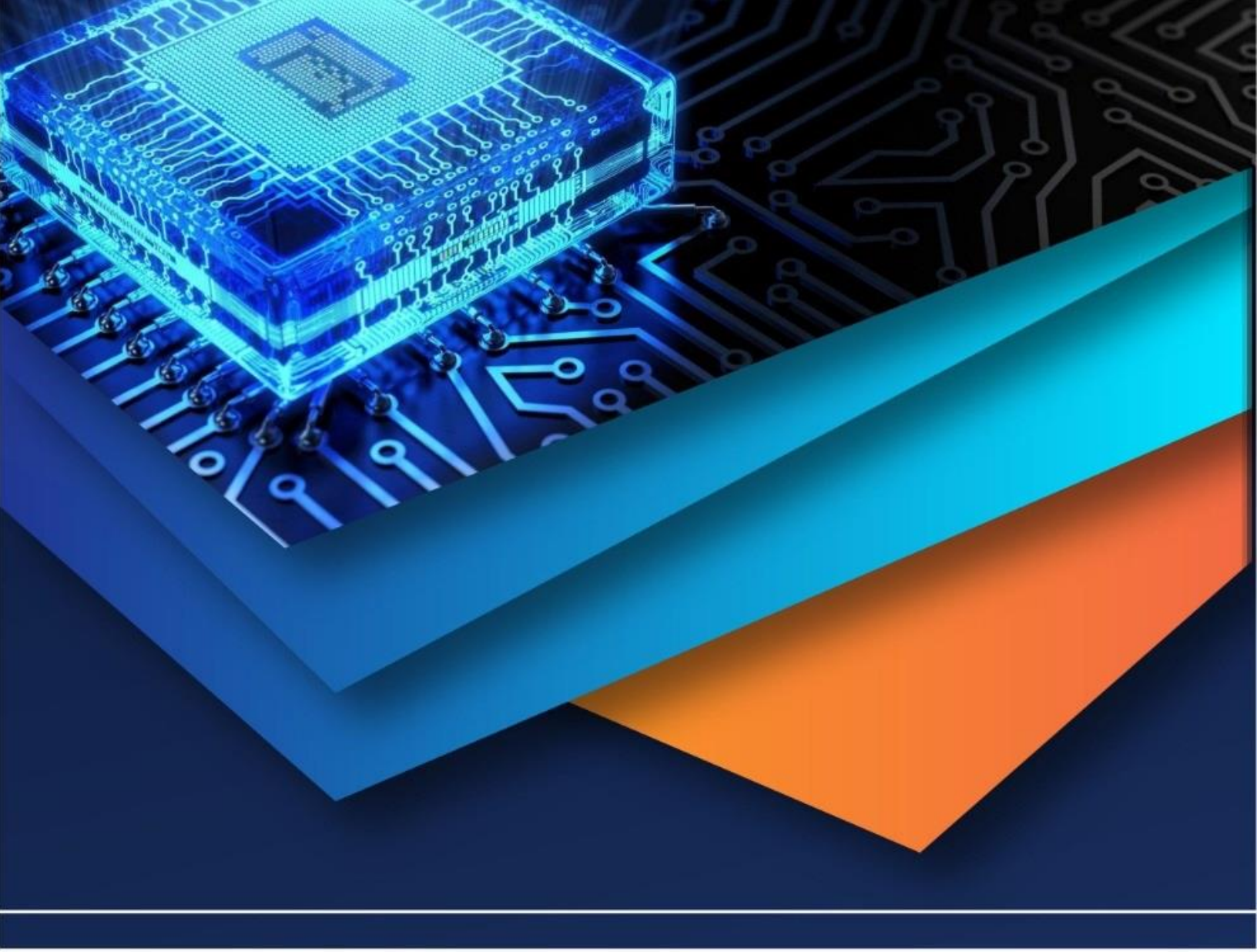
Python is chosen for web scraping because it offers convenient modules like 'urllib2', making it easy to access websites and extract information. MS Excel, on the other hand, is chosen for its comprehensive spreadsheet capabilities, supporting various file extensions, and providing a wide range of features, including data visualization and VBA for custom functions. Both Python and MS Excel complement each other in this project, with Python used for data collection and analysis, and MS Excel used for data visualization and cleaning tasks.

VII. CONCLUSION

This document presents a tale approach for credit card cheating perception using a hybrid of supervised and unsupervised learning techniques. The proposed method involves data decay based on Kernel Principal Component Analysis to project and decompose property varying for XGBoost, enhancing its ability to detect fraudulent behavior. Unlike previous methods, the personal of customers is taken into cogitation in this approach. By employing unsupervised learning techniques, the model is able to procedure the data unescorted by needing to know the meaning of every property, which addresses the challenge of property engineering. This approach not only enhances the fraud detection capabilities but also safeguards the privacy of clients by not explicitly accessing all features of the data. The use of XGBoost, a powerful gradient boosting algorithm, further improves the model's performance in detecting fraudulent transactions. The combination of supervised and unsupervised learning allows for better adaptability to changing transaction patterns and concept drift over time. Overall, the proposed hybrid method shows promising results in credit card fraud detection while considering client privacy. The approach demonstrates the importance of utilizing advanced machine learning techniques and addressing real-world challenges such as concept drift and imbalanced data in fraud detection systems. The findings of this research open up new possibilities for enhancing fraud detection systems and protecting both businesses and customers from financial losses due to credit card fraud.

REFERENCES

[1] Andrew. Y. Ng, Michael. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes", Advances in neural information processing systems, vol. 2, pp. 841-848, 2002.
[2] A. Shen, R. Tong, Y. Deng, "Application of classification models on credit card fraud detection", Service Systems and Service Management 2007 International Conference, pp. 1-4, 2007.
[3] A. C. Bahnsen, A. Stojanovic, D. Aouada, B. Ottersten, "Cost sensitive credit card fraud detection using Bayes minimum risk", Machine Learning and Applications (ICMLA). 2013 12th International Conference, vol. 1, pp. 333-338, 2013.
[4] B.Meena, I.S.L.Sarwani, S.V.S.S.Lakshmi," Web Service mining and its techniques in Web Mining" IJAEGT,Volume 2,Issue 1, Page No.385-389.
[5] F. N. Ogwueleka, "Data Mining Application in Credit Card Fraud Detection System", Journal of Engineering Science and Technology, vol. 6, no. 3, pp. 311-322, 2011.
[6] G. Singh, R. Gupta, A. Rastogi, M. D. S. Chandel, A. Riyaz, "A Machine Learning Approach for Detection of Fraud based on SVM", International Journal of Scientific Engineering and Technology, vol. 1, no. 3, pp. 194-198, 2012, ISSN ISSN: 2277-1581.
[7] K. Chaudhary, B. Mallick, "Credit Card Fraud: The study of its impact and detection techniques", International Journal of Computer Science and Network (IJCSN), vol. 1, no. 4, pp. 31-35, 2012, ISSN ISSN: 2277-5420.
[8] M. J. Islam, Q. M. J. Wu, M. Ahmadi, M. A. SidAhmed, "Investigating the Performance of Naive-Bayes Classifiers and KNearestNeighbor Classifiers", IEEE International Conference on Convergence Information Technology, pp. 1541-1546, 2007.
[9] R. Wheeler, S. Aitken, "Multiple algorithms for fraud detection" in Knowledge-Based Systems, Elsevier, vol. 13, no. 2, pp. 93-99, 2000.
[10] S. Patil, H. Somavanshi, J. Gaikwad, A. Deshmane, R. Badgujar, "Credit Card Fraud Detection Using Decision Tree Induction Algorithm", International Journal of Computer Science and Mobile Computing (IJCSMC), vol. 4, no. 4, pp. 92-95, 2015, ISSN ISSN: 2320-088X.
[11] S. Maes, K. Tuyls, B. Vanschoenwinkel, B. Manderick, "Credit card fraud detection using Bayesian and neural networks", Proceedings of the 1st international naiso congress on neuro fuzzy technologies, pp. 261-270, 2002.
[12] S. Bhattacharyya, S. Jha, K. Tharakunnel, J. C. Westland, "Data mining for credit card fraud: A comparative study", Decision Support Systems, vol. 50, no. 3, pp. 602-613, 2011.
[13] Y. Sahin, E. Duman, "Detecting credit card fraud by ANN and logistic regression", Innovations in Intelligent Systems and Applications (INISTA) 2011 International Symposium, pp. 315-319, 2011.
[14] Selvani Deepthi Kavila,LAKSHMI S.V.S.S.,RAJESH B " Automated Essay Scoring using Feature Extraction Method " IJCER , volume 7,issue 4(L), Page No. 12161-12165.
[15] S.V.S.S. Lakshmi, K.S.Deepthi, Ch.Suresh "Text Summarization basing on Font and Cue-phrase



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY