International Journal of Innovative Research in Engineering
Volume 5, Issue 2 (March-April 2024), PP: 165-171.
https://www.doi.org/10.59256/ijire.20240502022
www.theijire.com
ISSN No: 2582-8746

A Comprehensive Approach to Safeguard Credit Card Transactions and Fraud Prevention

Manish Kumar[1], Sushma Kumari[2], Rinku Kumar[3], Ajit Kumar[4], Somnath Banerjee[5], Kaustuv Bhattacharjee[6], Anirban Das[7]
[1,2,3,4,5,6,7] *Department of Computer Application, University of Engineering and Management, Kolkata, West Bengal, India.*

How to cite this paper:
Manish Kumar[1], Sushma Kumari[2], Rinku Kumar[3], Ajit Kumar[4], Somnath Banerjee[5], Kaustuv Bhattacharjee[6], Anirban Das[7], "A Comprehensive Approach to Safeguard Credit Card Transactions and Fraud Prevention", IJIRE-V5I02-165-171.

Abstract: *The escalating prevalence of financial fraud within the financial sector poses profound challenges. Detecting credit card fraud in online transactions necessitates data mining due to inherent complexities. Addressing two key issues—evolving patterns in legitimate and fraudulent behaviors and highly skewed datasets of credit card frauds—renders the task challenging. This paper scrutinizes the performance of naive Bayes, KNN, and logistic regression on significantly imbalanced credit card fraud data comprising 284,807 transactions from European cardholders. The dataset's skewness is addressed through a hybrid under-sampling and oversampling approach. The three techniques are applied to both unprocessed and preprocessed data.*

Fraud detection, defined as a set of activities thwarting illicit acquisition of assets or funds through deceptive means, varies across industries and methods. Credit card fraud, particularly susceptible due to its ease and prevalence in e-commerce and online platforms, prompted the adoption of diverse machine learning strategies to combat rising fraud rates. This paper employs machine learning algorithms for credit card fraud detection, utilizing a publicly available credit card dataset for model evaluation. While acknowledging that achieving 100% accuracy in fraud detection is elusive, the paper emphasizes the real-world applicability of its findings through the analysis of credit card data from a financial institution.

In addition to assessing model efficacy, the study introduces noise into the data samples to evaluate algorithm robustness. Experimental outcomes underscore the effectiveness of the majority voting method, achieving commendable accuracy rates in detecting credit card fraud cases. The study sheds light on the pressing issue of credit card fraud, emphasizing the importance of deploying robust machine learning approaches for timely and accurate detection in real-world scenarios.

Keyword: *Types of Fraud, How Do Fraudsters, Machine Learning, Credit Card fraud, Future Scope, Identify Thef, Fraudulent Transection, Cybercrime, Phishing,*

I.INTRODUCTION

Fraud is a deceptive act intending to conceal the truth and lead others to harmful actions. It is an offense that affects many individuals, either through personal experiences or stories shared by friends. Identity theft, a common form of fraud, involves the misappropriation of personal information such as bank account and credit card details. Many victims of identity theft have suffered significant financial losses and damage to their credit ratings.

Fraud is a serious crime that should not be taken lightly. It is punishable in most jurisdictions, with varying degrees of severity for different types of fraud.

Types of Fraud

While there are two main categories of fraud, several subtypes exist. Here are some of the most prevalent forms:
* Fake invoices, where fraudsters deceive others through fraudulent billing documents.
* Insurance fraud, which involves making false claims for financial gain.
* Tax evasion, where individuals misrepresent their income to avoid paying taxes.
* Check fraud, in which counterfeit or unauthorized checks are used for illicit purposes.
* Fraudulent transactions on the internet, including the sale of counterfeit goods.
* Fake websites designed to deceive visitors.
* Charity fraud, where organizations fail to allocate funds to the intended beneficiaries.
* Pyramid schemes, in which individuals are enticed to join and recruit others for financial gain.
* Work-from-home scams, which trick individuals into paying for information on earning opportunities.
* Credit card fraud, which involves unauthorized use of someone's credit card for monetary gain.

Credit Card Fraud

Credit card fraud is a significant concern in our technologically advanced world. Criminals employ various methods to commit credit card theft, including stealing cards before they reach their owners, obtaining card details through fraudulent

phone calls, exploiting mobile phone vulnerabilities, and stealing lost or misplaced cards.
Online transactions, where credit card information is easily used, have further exacerbated this problem. In 2017 alone, illegal card operations caused the deaths of 16.7 million individuals.

Key Concerns

As the prevalence of credit card fraud continues to rise, several key concerns arise, feasible to track down the individuals involved in credit card fraud and determine their current location, Can victims easily obtain refunds for fraudulent credit card transactions, How are ordinary people affected by technological advancements that facilitate credit card fraud.

II.BACKGROUND STUDY

In today's age of digital convenience, credit card fraud has become increasingly prevalent. It occurs due to a combination of carelessness with personal data and security breaches on websites. Understanding the various tactics employed by fraudsters can help safeguard our financial well-being. Let's explore some common methods employed by these cunning individuals:

One technique utilized by fraudsters is known as cloning. This involves using sophisticated software to duplicate a credit card's details and transferring them onto another card. This replication process is aptly referred to as skimming. By seamlessly transferring the data, fraudsters can make unauthorized transactions and gain access to someone else's funds.

Smishing: A Text Message Trap

Another strategy employed by scammers is smishing. This devious method relies on the trust we place in our mobile devices. Scammers send text messages containing deceptive URLs, persuading us to click on them. But beware! Clicking on such links can result in the download of malicious material onto our smartphones. This, in turn, allows fraudsters to extract all our personal information stored on the device, putting us at risk of credit card fraud.

Vishing: A Disguise of Trust

Fraudsters are known for their deceitful tactics, and vishing is no exception. In this scheme, they impersonate bank employees and call unsuspecting cardholders, pretending to be concerned about an issue with their account. They cleverly extract sensitive information such as card numbers, CVV codes, OTPs, and expiration dates. However, it's important to remember that reputable banks never request this kind of personal information over the phone. Stay alert and never disclose such details to unsolicited callers.

Identity Theft: A Web of Deception

Identity theft is a harrowing crime where fraudsters create fictitious identities that closely resemble real individuals. By adopting another person's identity, they gain access to their credit card information and exploit it for personal gain. A truly disconcerting act, but one that can be guarded against with vigilance and caution.

Now that we are aware of some common tactics employed by fraudsters, let's explore essential precautions we can take to protect ourselves from credit card fraud:

Important Precautions to Prevent Credit Card Fraud

Never lend your card to anyone else: It may seem obvious, but sharing your card with others can lead to unwanted consequences. Keep it safe and secure in your possession.

Guard your banking information: Protect your sensitive banking information by refraining from entering it on any suspicious internet portals or social media sites. Your financial well-being should never be jeopardized by sharing such details online.

Avoid writing down card numbers or PINs: In today's digital era, there's no need to jot down your credit card number or PIN. Preserve their confidentiality by committing them to memory instead.

Be cautious when disclosing card numbers or OTPs over the phone: Unless you initiated the call and are certain of the recipient's authenticity, exercise caution when divulging card numbers or OTPs over the phone. It's better to be safe than sorry.

Keep track of your card and receipts: It's important to stay mindful of your credit card's whereabouts and ensure you safely store any accompanying receipts. This will help you detect any discrepancies or potential issues.

Report loss or suspicious activity immediately: If you misplace your card, experience a change of address, or notice any suspicious activity on your account, contact your credit card company immediately. Prompt action can prevent further damage and ensure your financial security.

Literature Review

In the vast realm of literature, the study of fraud, the intentional act of seeking unauthorized financial gain, has been thoroughly explored. Researchers have dedicated their efforts to uncover various techniques for detecting anomalies and fraud, such as data mining applications, automated fraud detection systems, and even adversarial detection methods. Among the notable approaches are hybrid data mining and complex network classification algorithms, which have proven to be effective

in dealing with medium-sized online transaction datasets. A successful fraud detection system must have the ability to accurately identify instances of fraud swiftly while also avoiding the misclassification of legitimate transactions as fraudulent.

Proposed System

Our project has a primary objective: to raise awareness about Credit Card Fraud Detection and protect individuals from falling victim to online credit card fraud. We place great importance on the core focus of our system, which centers around securing transactions and ensuring utmost safety. Our ultimate aim is to detect 100% of fraudulent transactions while minimizing the occurrence of false classifications. Allow us to walk you through the functionalities of our proposed system through an illustrated block diagram that captures the essence of Credit Card Fraud Detection.

Dataset Description

Let's dive into the fascinating world of transaction datasets! In this dataset, we have an extensive collection of transactions made by a cardholder over a span of two days in September 2013. A whopping total of 284,807 transactions took place during this period.

Out of these numerous transactions, we discovered something quite intriguing. Prepare to be amazed—the dataset contains 492 fraudulent transactions! Yes, you read that right—a mere 0.172 percent of all transactions turned out to be fraudulent. Talk about an imbalance!

Confidentiality and Classification

To maintain the confidentiality of user transaction information, certain customer-related features have been classified as strictly confidential. We must respect the sensitivity of this data and preserve the privacy of these features.

Principal Component Analysis (PCA)

To bring some order to the chaos of features, we employed a powerful technique called Principal Component Analysis (PCA). This transformation involves applying PCA to features labelled as V1, V2, V3,..., V28, as well as the 'Time' feature. But hold on! Not all features undergo this transformation. 'Amount' and 'Class' remain non-PCA applied characteristics. Fascinating, isn't it?

III.METHODOLOGY

Detect Credit Card Fraud with Machine Learning

A dedicated data science team analyzes your data and creates models to identify and prevent illegal or fraudulent transactions.

This involves integrating various aspects of cardholder transactions, such as date, user location, product category, amount, supplier and customer behavior patterns.

The data is then fed into a machine learning model that is trained to recognize patterns and rules to determine the likelihood of fraudulent transactions.

Methods & Techniques Used

The following methods and techniques are utilized in the fraud detection process:

- Logistic Regression
- Decision Trees
- Random Forest
- KNN
- Isolation Forest
- Local Outlier Factor

We use a combination of supervised and semi-supervised machine learning techniques to effectively detect fraud. This approach overcomes several challenges, including severe class imbalance, inclusion of labeled and unlabeled samples, and handling a large number of transactions.

Fraud detection is a set of measures aimed at preventing money or property from being obtained under false pretenses. Various fraud detection datasets are used to provide a comprehensive view of both legitimate and illegitimate payment information. The detection system considers various factors such as IP address, geolocation, device ID, “BIN” data, global latitude/longitude, transaction history patterns, and actual transaction records. Merchants and publishers use business rules or analytical algorithms to analyze both internal and external data to detect fraud.

Machine Learning Techniques

Machine learning often categorizes tasks based on how data is collected and how the system responds to the data.

Two widely used machine learning techniques are unsupervised learning, which uses algorithms without labeled data to find structure in input data, and algorithms that use algorithms without labeled data to find structure in input data, and algorithms based on labeled input and output data provided by humans.

Tools & Technologies Used

Python: Python is a powerful programming language that offers a wide range of capabilities. It is known for its support of meta-programming and meta-objects and is fully compatible with object-oriented programming. Python also integrates power

typing, reference computation, and waste management, making it a versatile language for developers. With its advanced word processing capabilities, such as late binding, Python allows for the implementation of additional paradigms like contract generation and logic programming through extensions. To improve speed, developers can utilize mod-written modules in C-languages or Py for time-sensitive tasks. Python's architecture also incorporates features reminiscent of Lisp culture, including filters, maps, job reduction, list comprehensions, dictionaries, sets, and generator expressions.

Machine Learning

Machine Learning plays a crucial role in determining the fraudulent nature of credit card transactions. As a branch of science, machine learning enables computers to learn without explicit programming. It relies on data and patterns to continuously improve its performance and is widely employed in various fields, including fraud detection services. In this specific case, the dataset consists of credit card transactions from European cardholders in September 2013, with 492 instances of fraud out of 284,807 transactions within the previous two days. This accounts for approximately 0.172%. To address privacy concerns, independent variables are transformed numerically using Principal Component Analysis (PCA).

Supervised Learning

In the world of machine learning, supervised learning is a fascinating approach that involves providing the machine with sample inputs and their corresponding labeled expected outcomes. It's like teaching a clever student by showing them examples and guiding them through the learning process. The algorithm used in supervised learning "learns" by comparing its actual output to the expected outputs, identifying any flaws or areas for improvement, and updating its model accordingly.

Think of it as a teacher grading a test. The algorithm looks at the expected answers and compares them to its own. By doing so, it learns the patterns and connections between the inputs and outputs. This newfound knowledge allows it to make predictions and generate label values for unlabeled data.

Let's say we train an algorithm using labeled data of sharks as fish and oceans as water. Now, with this knowledge in its arsenal, the algorithm becomes an expert at recognizing patterns and can easily identify unlabeled shark images as fish and ocean images as water. It's like a superpower that allows the algorithm to make sense of the world even when it's presented with data it has never seen before.

Unsupervised Learning

While supervised learning is all about guidance and labeled data, unsupervised learning takes a different approach. In this case, the learning algorithm is faced with a unique challenge - it must discover patterns and commonalities within a dataset that is completely unlabeled. It's like exploring a vast uncharted territory without any directions or signposts.

Without being given explicit labels, the algorithm has to rely on its innate learning abilities to discern meaning and connections within the data. It's like solving a puzzle without knowing what the final image looks like. This requires a keen eye for detail and an ability to detect hidden patterns or features that may not be immediately apparent.

Consider the example of a sophisticated unsupervised learning system analyzing customer transactions. Without any labels to guide it, the algorithm is still able to deduce remarkable insights. For instance, it might recognize that women of a certain age group who prefer unscented soaps are likely pregnant. This sort of information can be incredibly valuable for targeted marketing campaigns, allowing businesses to tailor their efforts and create more meaningful connections with their customers.

Hardware Requirements:

- Processors: The system should have a Pentium 4, Intel Core i3, i5, or i7 processor with a minimum speed of 2 GHz. This ensures efficient processing of data.
- RAM Capacity: The system must have a minimum of 512MB RAM to handle the computational tasks effectively.
- Hard Disk Capacity: A minimum of 10 GB disk space is necessary to install and store the required software and data files.
- Input Devices: A keyboard and mouse are required for user interaction with the system.
- Output Devices: A monitor or PC is needed to display the output and enable the user to visualize the results.

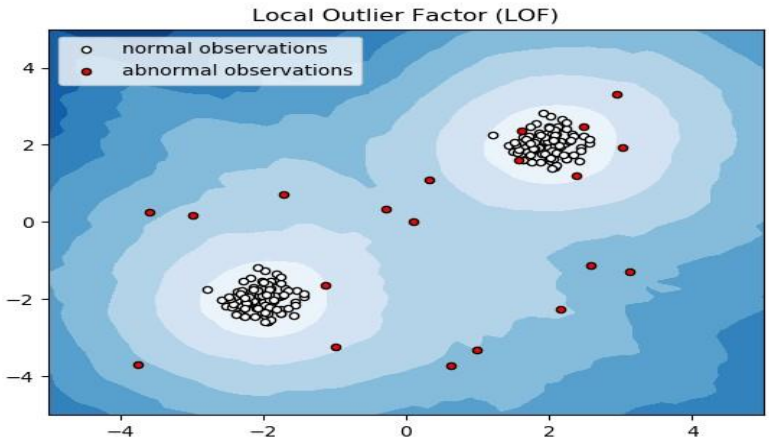
Software Requirements:

- Supported Operating Systems: The proposed methodology is compatible with Windows 7, 10, and above. These operating systems provide a stable and secure environment for the software to run smoothly.
- Platforms: The preferred platform for implementing the proposed methodology is Jupyter Notebook. Jupyter Notebook offers a user-friendly interface and supports interactive data exploration and analysis.
- Programming Language: Python is the recommended programming language for implementing the methodology. Python is widely used and offers a rich set of libraries and tools for machine learning and data analysis.
- Tools: To effectively implement the proposed methodology, the following tools are required: Python Django, PostgreSQL, and relevant files. These tools provide the necessary functionalities and resources for performing anomaly detection.

IV. RESULTS AND DISCUSSION

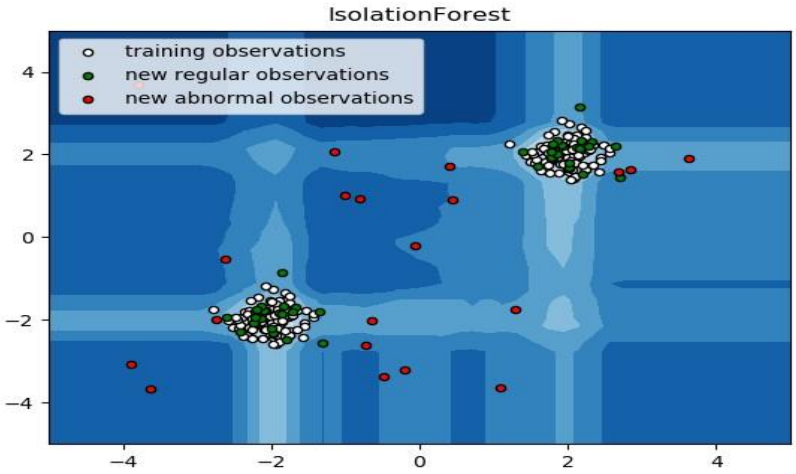
Implementing this concept is difficult because it requires market competition, legal considerations, and cooperation with banks that are reluctant to share data to protect user information. Therefore, we consulted references that investigated similar approaches and collected results. One of these references states: "This approach was applied to a comprehensive application dataset provided by a German bank in 2006. Reasons for Bank Secrecy." Below we provide only a summary of the

results achieved. Using this method, the Level 1 list will contain some cases that are most likely to be fraud. Everyone on this list is at risk and their cards are blocked to avoid risk. For other lists, the situation is more complicated. Level 2 listings remain limited and must be considered on a case-by-case basis. Credit reporting and collection agencies estimate that half of the cases on this list may be suspected fraud. For the remaining largest list, the work is quite extensive. Less than a third are suspect. One way to maximize time efficiency and reduce overhead is to include new elements in your queries. For example, this element can be a phone number, email address, or the first five digits of a password. These new queries can be applied to level 2 and 3 lists. Upon visualizing the result of the Local Outlier Factor algorithm



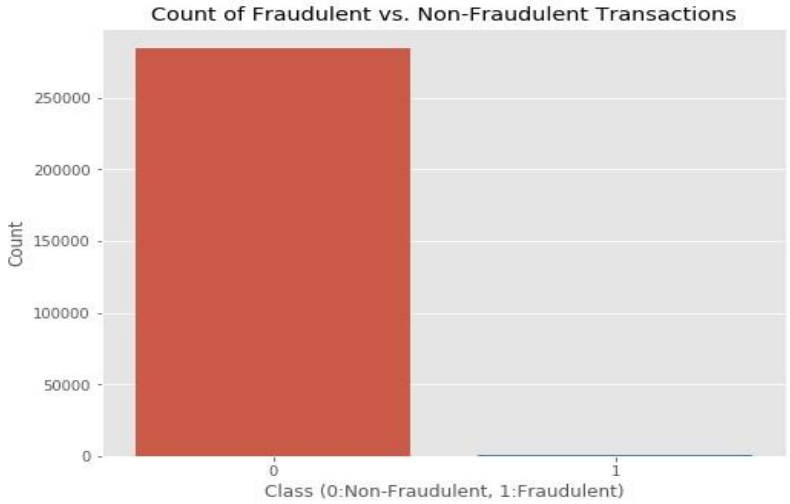
By comparing the local values of a sample with those of its neighbours, it is possible to identify samples that are significantly lower than their neighbours. These values are considered as outliers. Due to the large size of the dataset, we conducted our tests using only a fraction of it to reduce processing times. The final result with the complete dataset processed is also determined and is presented in the results section of this paper.

On plotting the results of Isolation Forest algorithm

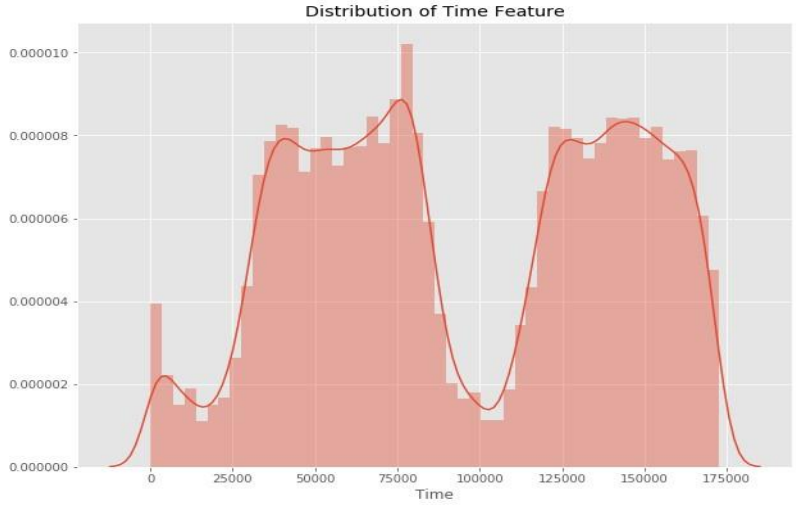


Randomly partitioning them results in shorter paths for anomalies. When a forest of random trees simultaneously generates shorter path lengths for specific samples, they are highly likely to be anomalies. Once anomalies are detected, the system can be utilized to report them to the relevant authorities. For testing purposes, we are evaluating the outputs of these algorithms to determine their accuracy and precision.

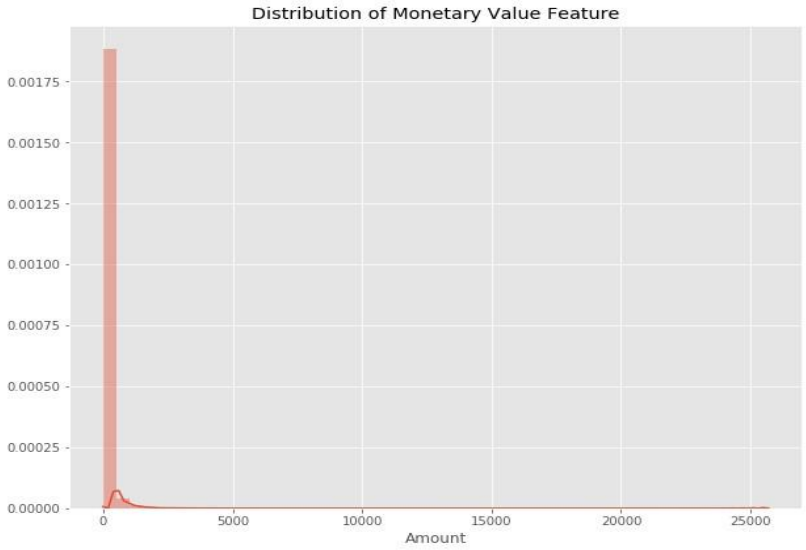
Initially, the dataset was sourced from Kaggle, a website dedicated to data analysis and providing datasets. Within this dataset, there are 31 columns, and to safeguard sensitive information, 28 of them are denoted as v1-v28. The remaining columns include Time, Amount, and Class. The Time column indicates the time gap between the initial transaction and the subsequent one, while Amount represents the monetary value of the transaction. Class 0 designates a legitimate transaction, and 1 denotes a fraudulent one. To scrutinize the dataset for irregularities and gain a visual understanding, various graphs were plotted.



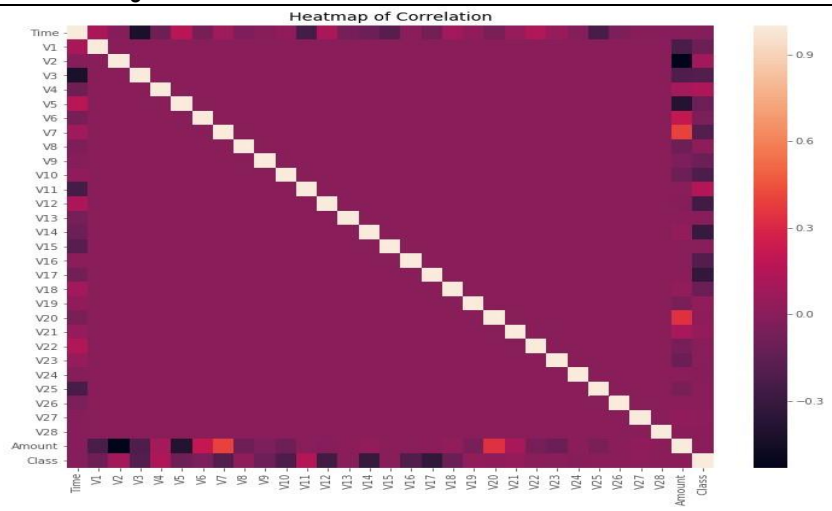
This graph shows that the number of fraudulent transactions is much lower than the legitimate ones



This chart illustrates the times at which transactions were completed within a day. It is evident that the fewest transactions occurred during the nighttime and peaked during the daytime.



The heat map is displayed below:



V.CONCLUSION

Credit card fraud represents a form of criminal deception, and this article has delineated the prevalent fraud techniques, along with their detection strategies. It has also examined recent advancements in this domain. The paper provides a detailed exploration of how machine learning can enhance fraud detection, elucidating the algorithms, pseudocode, implementation rationale, and experimental outcomes.

While the algorithm attains an accuracy exceeding 99.6%, its precision remains at 28% when considering only a tenth of the dataset. However, with the complete dataset inputted into the algorithm, precision rises to 33%. This high accuracy percentage is expected due to the significant imbalance between the number of legitimate and fraudulent transactions.

As the entire dataset covers only two days' transaction data, it represents a fraction of the information that could be available on a commercial scale. The system's efficiency is anticipated to improve over time as more data is incorporated, leveraging the benefits of machine learning algorithms.

Future Scope

Although our goal of achieving 100% accuracy in fraud detection wasn't reached, the developed system shows promise and can approach that goal with sufficient time and data. This project inherently allows for the integration of multiple algorithms as modules, with their results combined to enhance the final accuracy. The model's flexibility and modularity are evident in the ease of adding new algorithms, as demonstrated in the code.

Further improvements can be explored by incorporating additional algorithms into the model, provided their output adheres to the same format. This feature enhances the project's modularity and flexibility. Additional opportunities for enhancement lie in the dataset; as demonstrated earlier, increasing the dataset size improves algorithm precision. However, achieving this requires genuine support from financial institutions.

References

1. *"Credit Card Fraud Detection Based on Transaction Behaviour -by John Richard D. Kho, Larry A. Vea" published by Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017*
2. *CLIFTON PHUA1, VINCENT LEE1, KATE SMITH1 & ROSS GAYLER2 " A Comprehensive Survey of Data Mining-based Fraud Detection Research" published by School of Business Systems, Faculty of Information Technology, Monash University, Wellington Road, Clayton, Victoria 3800, Australia*
3. *"Survey Paper on Credit Card Fraud Detection by Suman" , Research Scholar, GJUS&T Hisar HCE, Sonepat published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2014*
4. *"Research on Credit Card Fraud Detection Model Based on Distance Sum – by Wen-Fang YU and Na Wang" published by 2009 International Joint Conference on Artificial Intelligence*
5. *"Credit Card Fraud Detection through Parenclitic Network Analysis By Massimiliano Zanin, Miguel Romance, Regino Criado, and SantiagoMoral" published by Hindawi Complexity Volume 2018, Article ID 5764370, 9 pages*
6. *"Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy" published by IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 29, NO. 8, AUGUST 2018*
7. *"Credit Card Fraud Detection-by Ishu Trivedi, Monika, Mrigya, Mridushi" published by International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 1, January 2016*
8. *David J.Wetson,David J.Hand,M Adams,Whitrow and Piotr Jusczak "Plastic Card Fraud Detection using Peer Group Analysis" Springer, Issue 2008.*