CREDIT CARD FRAUD DETECTION

A major project report submitted in partial fulfillment of the requirement for the award of degree of

Bachelor of Technology
in
Computer Science & Engineering / Information Technology

Submitted by

Saksham Sharma (201382)

Aryan Bhardwaj (201479)

Under the guidance & supervision of

Dr. Pradeep Kumar

Department of Computer Science & Engineering and Information Technology
Jaypee University of Information Technology, Waknaghat, Solan - 173234 (India)

CERTIFICATE

This is to certify that the work which is being presented in the project report titled "CREDIT CARD FRAUD DETECTION" in partial fulfilment of the requirements for the award of the degree of B. Tech in Computer Science And Engineering and submitted to the Department of Computer Science And Engineering, Jaypee University of Information Technology, Waknaghat is an authentic record of work carried out by Aryan Bhardwaj(201479),Saksham Sharma (201382) during the period from July 2023 to June 2024 under the supervision of Dr.Pradeep Kumar, ,Department of Computer Science and Engineering, Jaypee University of Information Technology, Waknaghat.

Name: Aryan Bhardwaj
Enrollment No.: 201479

Name: Saksham Sharma
Enrollment No.: 201382

The above statement made is correct to the best of my knowledge.

Dr. Pardeep Kumar
Associate Professor,SM-ACM
Jaypee University of Information Technology, Waknaghat

Candidate's Declaration

I hereby declare that the work presented in this report entitled 'CREDIT CARD FRAUD DETECTION' in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering / Information Technologysubmitted in the Department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology, Waknaghat is an authentic record of my own work carried out over a period from July 2023 to June 2024 under thesupervision of Dr. Pradeep Kumar.

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

(Student Signature with Date)
Name: Aryan Bhardwaj
Enrollment No.: 201479

(Student Signature with Date)
Name: Saksham Sharma
Enrollment No.: 201382

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

(Supervisor Signature with Date)
Supervisor Name: Dr. Pradeep Kumar
Designation: Associate Professor, SM-ACM
Jaypee University of Information Technology, Waknaghat
Dated:

ACKNOWLEDGEMENT

It gives me an incredible feeling of joy the report of the B .Tech Minor Project attempted during B .Tech 4 th year. This venture is itself is an affirmation to the motivation, drive and specialized help added to it by numerous people. This venture could never have seen the light of the day without the assistance and direction that we have gotten. furnishing me with an urging stage to build up this venture, which in this manner helped me in molding my capacities towards a valuable objective. of C.S.E & I.T, for his steady help and direction over the span of our work. His sincerity, exhaustiveness and presistance have been a consistent source of enthusiasm for me. He has given all of his widely experienced thoughts and wise remarks at for all intents and purposes all phases of the venture and has likewise shown us the most recent industry-situated advancements. I moreover couldn't care less to miss the opportunity to perceive the responsibility of all representatives of the division for their caring direction or collaboration during the progress of my challenge. Last yet not the least, we perceive our allies for their responsibility in the completion of the endeavor. My heartiest gratitude to Dr. Vivek Sehgal, Head of Dept., Department of C.S.E & I.T for I owe uncommon obligation. Of appreciation to Dr.Pradeep Kumar, Associate Professor Department.

Name : Saksham Sharma

Name: Aryan Bhardwaj

TABLE OF CONTENTS

CHAPTER-7: CONCLUSION AND FORESEEABLE ENCHANCEMENTS

REFERENCES

APPENDIX

ABSTRACT

Due to the rapid growth of E-Commerce, the use of credit cards for online transactions has significantly increased, leading to a rise in credit card fraud. As credit cards become the preferred method of payment for both online and traditional transactions, instances of fraud associated with them have also risen. Traditional fraud detection methods often fall short in accurately identifying fraudulent transactions, as fake transactions are blended with legitimate ones.

To mitigate losses, credit card issuers find it imperative to implement effective fraud detection systems. Advanced techniques based on Artificial Intelligence, Data mining, Fuzzy logic, Machine learning, Sequence Alignment, and Light GBM software programming have emerged to identify various fraudulent credit card transactions. A comprehensive understanding of these methods is crucial in developing a successful credit card fraud detection system.

This paper provides a review of different approaches employed in credit card fraud detection tools and assesses each method based on specific criteria. The goal is to establish a reliable and efficient system for detecting and preventing credit card fraud. Additionally, continuous research and innovation are essential in staying ahead of evolving fraud tactics, ensuring the ongoing security of credit card transactions in the dynamic landscape of E-Commerce.

Furthermore, it is essential for consumers to stay informed about security measures and be vigilant in monitoring their credit card statements for any suspicious activities. Education and awareness campaigns can play a vital role in empowering individuals to recognize potential threats and take proactive steps to safeguard their financial information. Collaboration between financial institutions, businesses, and regulatory bodies is crucial in fostering a secure and trustworthy environment for electronic transactions. As the digital landscape continues to evolve, the commitment to enhancing security measures must remain a priority to build and maintain trust in the use of credit cards for transactions.

CHAPTER 1
INTRODUCTION

WHAT IS A CREDIT CARD?

- A credit card issued by a credit card provider, such as Capital One, is designed for making purchases in stores or online.
- Credit cards are commonly used for balance transfers and withdrawing cash from ATMs.
- These cards can be used globally, being accepted at numerous locations, though there may be fees associated with international use.

Upon receiving a credit card, you are assigned a credit limit, indicating the total amount available for spending using the card. It is advisable to maintain some available credit on your card to account for any interest that may be applied.

All transactions you make will be reflected on your statement, including details such as the minimum payment required and the deadline by which your credit card provider must receive at least the minimum payment.

If you do not pay off your balance in full each month, interest will be charged on the remaining amount. It is crucial to review your credit card agreement to understand the applicable charges and fees.

In addition to facilitating transactions, credit cards are commonly used for balance transfers and cash withdrawals from ATMs. The global acceptance of credit cards has expanded, making them a

convenient means of payment in a multitude of locations. However, it's important to note that there may be fees or charges associated with using a credit card for transactions conducted abroad Upon obtaining a credit card, the cardholder is assigned a credit limit, which represents the maximum amount they can spend using the card. Responsible financial management involves leaving a portion of this credit limit available to account for potential interest charges.

Details of all transactions made with the credit card are recorded on the monthly statement, providing a comprehensive overview of the cardholder's spending activities. The statement includes crucial information such as the minimum amount due, the deadline by which the credit card provider must receive at least the minimum payment, and any applicable charges.

It is imperative for credit card users to be aware of their credit card agreement, understanding the terms, conditions, and potential fees associated with the card. Failure to pay the balance in full each month may result in interest charges being applied to the remaining balance. Regularly checking the credit card agreement and statement ensures that cardholders are informed about the financial implications of their card usage, promoting responsible and informed financial practices.

.

WHAT ARE FRAUDULENT TRANSACTIONS?

The objective may involve acquiring goods without making payment or securing unauthorized funds from an account. Fraudulent transactions refer to requests and purchases conducted using a credit card or bank account not belonging to the legitimate cardholder or account holder.

This type of activity, a significant contributor to identity theft, has the potential to inflict harm on both merchants and the victim of identity fraud. Mitigating the impact of fraudulent transactions is crucial for the well-being of both sellers and buyers.

Therefore, it is essential for individuals to take precautionary measures when managing financial accounts to prevent such unauthorized activities andsafeguard against potential risks associated withfraudulent exchanges.

WHAT IS FRAUD DETECTION?

Detecting fraud involves an intricate process of monitoring and analyzing customer behavior to evaluate, identify, and prevent illicit activities. In the realm of credit card transactions, countering fraud is imperative for the security and integrity of financial systems. A nuanced comprehension of the evolving techniques employed by fraudsters is essential in developing effective strategies to safeguard against unauthorized activities.

Credit card security hinges not only on the physical integrity of the card itself but also on safeguarding the associated credit card number. The global surge in online transactions, propelled by the widespread use of the internet, has led to an unprecedented increase in credit card usage worldwide. This surge has necessitated a rapid development in the volume of credit card transactions,underlining the critical importance of robust fraud detection mechanisms.

Credit card fraud, a broad and encompassing term, involves theft and deception where a credit card serves as a false source of funds in each transaction. Fraudsters continually employ a myriad of sophisticated techniques to exploit vulnerabilities in the system. To effectively combat credit card fraud, it is not only vital to comprehend the underlying factors contributing to fraudulent activities but also to stay abreast of emerging trends and tactics.

Preventing credit card fraud requires a multi-faceted approach that integrates advanced technological solutions such as Artificial Intelligence, Data Mining, and Machine Learning. These tools empower financial institutions to analyze patterns, detect anomalies, and respond swiftly to potential threats. Additionally, fostering collaboration among financial institutions, regulatory bodies, and law enforcement agencies is crucial in creating a robust network to combat the global of credit card fraud.

The evolution and stabilization of credit card fraud prevention measures over the years underscore the resilience of the financial industry in adapting to emerging challenges. Continuous research, innovation, and a proactive stance are paramount in staying one step ahead of fraudsters and ensuring the ongoing security of credit card transactions. As we navigate the ever-changing landscape of financial technology, the commitment to enhancing security measures remains pivotal for maintaining trust and confidence in electronic transactions.

CHAPTER 2

PROBLEM DEFINITION AND FEASIBILITY ANALYSIS

2.1 PROBLEM DEFINITION

The aim is to develop a robust credit card fraud detection system utilizing the Light GBM algorithm. During credit card transactions, instances of fraud are identified, and the number of falsealarms is minimized by employing the Light GBM algorithm. To enhance the accuracy of fraud detection, we introduce a task function that considers the asymmetry in misclassification costs, thereby optimizing for correctness.

Light GBM distinguishes itself by constructing trees vertically, a departure from other algorithms that build trees uniformly. Specifically, Light GBM employs a leaf-wise approach, selecting the leaf with the maximum delta loss for creation.

This strategy can result in a more effective reduction of loss compared to level-wise algorithms when developing similar leaves.

As the volume of data continues to grow, traditional data science algorithms face challenges in delivering speedy results. Light GBM earns its 'Light' prefix due to its efficiency, efficiently managing large datasets and offering quicker results.

Moreover, it addresses the precision of outcomes, a key reason for its popularity. Light GBM's ability to handle vast amounts of data with lower computational overhead makes it a preferred choice for data scientists.

Notably, Light GBM supports GPU acceleration, contributing to its widespread adoption in data science applications. This feature aligns with the increasing demand for efficient and powerful tools in the rapidly evolving field of data science.

The emphasis on precision, speed, and adaptability positions Light GBM as a reliable and efficient solution for developing credit card fraud detection systems and other data science applications.

2.2 LITERATURE SURVEY

Light GBM earns its 'Light' prefix due to its remarkable speed. It efficiently manages vast amounts of data, requiring lower memory to operate. Another notable aspect contributing to LightGBM's popularity is its focus on result accuracy. Additionally, Light GBM supports GPU acceleration,
making it a preferred choice among data scientists for various applications in the field of data science.

Given the inherent nature of credit card transactions, a more nuanced and appropriate measure is imperative for fraud detection. The conventional approach, which focuses on the sheer quantity of accurately classified transactions, may fall short in addressing the unique challenges posed by credit card fraud.

A more relevant metric could involve limiting the total credit card unlock capacity
based on potentially fraudulent activity. This would be particularly useful in cases where a stolen orlost card is exploited by fraudsters until its available credit limit is exhausted.

Traditionally, fraud detection has been treated as a classification problem, with various statistical and data mining algorithms proposed to address it.

Decision trees and artificial neural networks have emerged as popular choices in this domain. Bolton and Hand's research provides a comprehensive overview of literature on fraud detection issues.

However, when approached as a classification problem with imbalanced misclassification costs, traditional data mining algorithms may not be directly applicable. Adjustments or new algorithms tailored specifically for this context may be necessary.

An alternative strategy could involve exploring the use of general-purpose metaheuristic techniques, such as the Light GBM algorithm, to address the unique challenges posed by credit card fraud detection. This adaptive approach reflects the dynamic and evolving landscape of fraud detection in the realm of credit card transactions.

2.2.1 LIGHTGBM ALGORITHM

Light GBM algorithms are innovative computational methods designed to secure improved solutions as circumstances evolve. Since their inception in Holland, they have proven highly effective in tackling diverse and challenging domains, particularly when confronted with massive datasets.

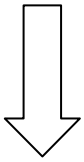
Notably, these algorithms have found applications in information mining, primarily for variable selection, and are frequently integrated with other data mining techniques. In the context of this study, our aim is to address the classification issue using a LightGBM algorithmic approach.

These evolving algorithms exhibit adaptability and have demonstrated their efficacy across various fields, showcasing their prowess in handling complex problems. Originating in Holland, LightGBM algorithms have evolved into versatile tools capable of processing large datasets efficiently. Their versatility extends to information mining, where they are often employed for tasks such as variable selection. Moreover, they seamlessly integrate with other cutting-edge data mining algorithms.

I the pursuit of solutions to our classification problem, we opt for a Light GBM algorithmic approach. This decision is rooted in the algorithm's ability to navigate the intricacies of evolving circumstances and its proven success in various challenging domains. By leveraging the strengths of Light GBM, we aim to enhance our ability to address the intricacies of our classification issue, ensuring a robust and adaptive solution.

[1,2]

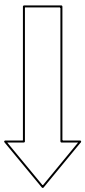
PSEUDO CODE OF LIGHT GBM ALGORITHM



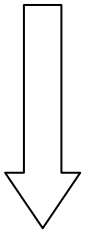
Introduce the information

Assess introductory

Information



Rehash Perform serious



Apply Light GBM administrators to create new

arrangements,evaluate arrangements in the information until

some convergence criteria is satisfied.

SELECTION PROCESS

Max_depth: This parameter defines the maximum depth of a tree, serving as a measure to control overfitting in the model. If you observe signs of overfitting in your model, a recommended approachis to decrease the max depth. This adjustment helps prevent the tree from becoming too complex and tailored to the training data, thus enhancing generalization to new, unseen data.

Min_data_in_leaf: This parameter sets the minimum number of records required in a leaf. Finding the optimal value for this parameter is crucial, as it aids in preventing overfitting by ensuring that each leaf contains enough data for meaningful decision-making. Adjusting this parametercontributes to achieving a balanced and well-generalized model.

Feature fraction: Particularly relevant during boosting (discussed later), this parameter is employed in random forests. Feature fraction dictates that Light GBM randomly selects a specified percentage of parameters in each cycle when constructing trees. This randomness aids in diversifying the trees, contributing to better model performance and robustness.

Bagging_fraction: This parameter determines the fraction of data utilized for each iteration during training. It is commonly employed to expedite the training process and avoid overfitting. By using a subset of the data for each iteration, the model can learn more efficiently while minimizing the risk of tailoring too closely to the training set.

Early stopping round: This parameter proves valuable in expediting model training. If, during the last early_stopping_round iterations, there is no improvement in a specific metric on a validation dataset, the training process will cease. This approach helps prevent excessive iterations, promoting efficiency in the model-building process and avoiding unnecessary computational resources. It's a practical tool for ensuring that the model achieves optimal performance without unnecessary computational overhead.

TOURNAMENT SELECTION

In this context, competition determination plays a pivotal role as it strategically selects the most optimal individuals from diverse groups. The mechanism involves the random selection of data, organizing them into competitions, and then determining the winner from each group. The victorious individual is subsequently placed into the mating pool for recombination. This iterative process is repeated until the desired size of intermediate data is achieved.

The competition amount serves as a critical parameter governing the quality of selection. Specifically, a larger competition size strengthens the selection process. The rationale behind this is that a larger pool of participants in each competition increases the competitive intensity, ensuring that only the most superior individuals proceed to the next generation. This method of competition-based selection contributes to the overall efficiency and effectiveness of the evolutionary algorithm, fostering the emergence of high-quality solutions over successive iterations.

LIMITED SELECTION

This is the most significant parameter and determines the utilization of your model, regardless of whether it is a relapse issue or order issue. Light GBM will of course consider model as a relapse model.

- regression: for relapse

- binary: for paired order

- multiclass: for multiclass characterization issue

LGBM WORKS

Consider a fundamental decision tree designed to predict whether an individual is likely to purchase a computer. In such a scenario, the model might predict that a young student is inclined to buy a computer, while an older individual with a less-than-stellar credit score may not be as likely to make that purchase.

While decision trees are renowned for their flexibility and interpretability, a solitary decision tree can be susceptible to overfitting, making it challenging to generalize effectively. Various techniques, such as constraining the tree's depth, have been employed to limit flexibility, but this often results in the tree underfitting the data.

To overcome this limitation, decision trees are commonly utilized
in combination rather than in isolation. Gradient Boosting Decision Trees (GBDTs) represent one method among many for amalgamating the predictions of multiple decision trees to create more robust and generalizable predictions.

The concept behind GBDTs is elegantly simple: merge the predictions of diverse decision trees by aggregating them. For instance, when predicting housing prices, the predicted cost for any given data point would be the sum of the predictions generated by each individual decision tree. This ensemble approach enhances the predictive power of decision trees, fostering more accurate and reliable predictions across a diverse range of scenarios.

The utilization of GBDTs exemplifies the synergy achieved by combining the strengths of multiple decision trees, providing a more nuanced and accurate model for predictive analysis.

EXECUTION

This generational procedure be rehashed awaiting an end situation has be reach Basic ending circumstances be:

- An arrangement are discovered which fulfills least models
- Set figure of ages came to
- Owed financial plan (calculation time/cash) arrived at the most elevated positioning arrangement's wellness is coming to or has arrived at a level with the end goal that progressive.

FEASIBILITY ANALYSIS:

An achievability investigation is a significant apparatus to assist you with surveying the reasonability of beginning another worth included business or re-sorting out or growing a current business.

All tasks are practical given boundless assets and unending time. Be that as it may, shockingly, shortage of assets and troublesome conveyance dates torment all ventures.

The following three kind of feasibilities are calculated in the probability investigation of the

project.

- Operational feasibility.
- Technical feasibility
- Economic feasibility.

TECHNICAL FEASIBILITY:

Specialized feasibility involves evaluating the technical viability of the envisioned system. The availability of essential hardware and software resources crucial for system development further enhances its technical feasibility. This phase scrutinizes the technical requirements of the proposed system and assesses the efficiency of the newly developed project in aligning with the existing technical prerequisites of the framework. As a result, the technical feasibility of the system is heightened, ensuring that it meets the necessary technical criteria for seamless functionality.

ECONOMIC FEASIBILITY:

Financial analysis, commonly referred to as cost/benefit analysis, stands as the predominant method for assessing the feasibility of a new system. This approach is widely adopted due to its effectiveness in weighing costs against benefits. In the context of this project, the software utilized is freeware, resulting in minimal development costs for the tool. The simplicity of the required procedure and the minimal programming involved contribute to its cost-effectiveness. Consequently, the project demands relatively low expenses and programming efforts, making it adaptable for deployment in diverse environments. The cost-benefit analysis, in this scenario, underscores the economic viability and accessibility of the system across various settings.

CHAPTER: 3

INTRODUCTION TO GENETIC ALGORITHM

An innate determination is an investigative pursuit inspired by Charles Darwin's theory of natural selection. This concept mirrors the mechanism of natural selection, wherein individuals with advantageous traits are chosen for reproduction, ultimately influencing the genetic makeup of the population over time.

NOTION OF ARTIFICIAL SELECTION

The model of natural selection mirrors an early developmental strategy that involves selecting appropriate individuals from a population. These chosen individuals then produce offspring inheriting their beneficial traits, contributing to the overall enhancement of the population. If parents exhibit superior fitness, their offspring are likely to surpass them and have an increased chance of long-term success. This iterative process continues, culminating in the emergence of future generations with increasingly robust and fit individuals.

Applying this model to problem-solving, genetic algorithms emulate the principles of natural selection. In the quest for optimal solutions, multiple iterations are executed, with the selection of the most suitable candidates reminiscent of the natural selection process. This approach involves evaluating and choosing the best solutions from a pool of candidates, much like the advantageous traits passed down through generations.

The genetic algorithm typically unfolds in five distinct stages:

1. Initial Population: A diverse set of potential solutions is generated to initiate the process.

2. Fitness Function: A metric, known as the fitness function, assesses the performance of each solution

in the context of the problem at hand.

3. Selection: Individuals that exhibit superior fitness are chosen to proceed to the next stage, simulating the natural selection of advantageous traits.

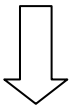
4. Crossover: Selected individuals exchange genetic material, combining their attributes to createpotentially superior offspring.

5. Mutation: Random changes are introduced to the genetic makeup of certain individuals, promotingdiversity and adaptability within the population.

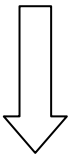
This iterative cycle continues, refining the solutions with each generation. The genetic algorithm, inspired by the principles of natural selection, stands as a powerful problem-solving tool, capable of efficiently navigating solution spaces to identify optimal outcomes. Through mimicking the mechanisms of natural selection, genetic algorithms showcase the potential for evolutionary-inspired approaches in addressing complex problems and optimizing solutions.

PSEUDO CODE:

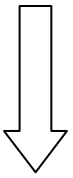
START



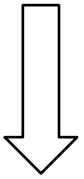
Produce the underlying populace



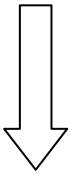
Compute Wellness work



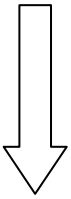
Repeat



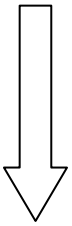
Starting Populance



Crossover



Mutation



Compute

wellness work

UNTIL population has converged

STOP

GENETIC ALGORITHM WORKS:

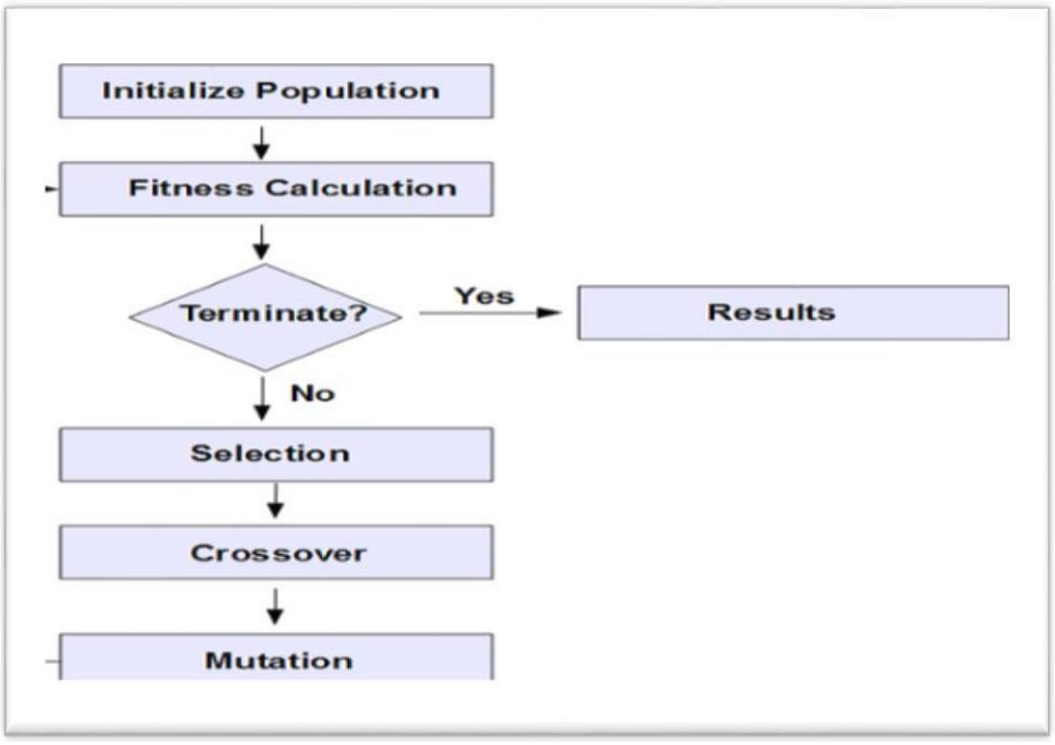


Figure: 2.1 Flow Diagram

CHAPTER 4

SOFTWARE REQUIREMENTS SPECIFICATION

INTRODUCTION .

The Software Requirements Specification (SRS) serves as a comprehensive documentation of a client's or potential client's organization understanding, outlining the system's specifications and conditions at a specific moment before the commencement of any actual design or development work. It plays a pivotal role in laying the groundwork for successful software projectsby providing a detailed roadmap for future reference, particularly in instances of uncertainty and misinterpretation.

The primary purpose of the Software Requirements Specification is to establish a clear and detailed outline of the system's essentials. This includes a meticulous delineation of practices, requirements, and the anticipated performance of the system. By encapsulating these critical elements, the SRS acts as a guiding document that not only directs the development team but also serves as a reference point for stakeholders, ensuring a shared understanding of the project's objectives and requirements.

REQUIREMENT ANALYSIS

The overarching goal is to comprehensively understand the client's requirements, employing a systematic approach that involves analyzing needs, assessing feasibility, devising a strategic plan, validating specifications, and effectively managing the evolving requirements throughout the software development life cycle.

PURPOSE

This document serves the purpose of delineating the essential requirements for the implementation of a credit card fraud detection system. In a comprehensive manner, it offers a holistic overview of our project, encompassing customer requirements, product perspective, an outline of requirements, and overarching limitations. Additionally, this document is designed to delve into specific requirements and functionalities critical for the successful execution of the project. These include aspects such as the user interface, functional prerequisites, and performance criteria.

The elucidation of customer requirements aims to encapsulate the expectations and needs of the end-users in the context of credit card fraud detection. It seeks to establish a clear understanding of the desired outcomes and functionalities from a user's perspective, laying the foundation for a system that aligns seamlessly with user expectations.

SCOPE

The scope of this Software Requirements Specification (SRS) report extends across the entire life cycle of the project. Its relevance is sustained from the project's initiation to its completion, providing a comprehensive reference point for the software development process. This document is instrumental in defining the ultimate state of the software requirements that have been mutually agreed upon by bothclients and designers. As a living document, it undergoes updates and refinements throughout the project's execution, reflecting the evolution of the software's features and functionalities.

This report serves as a guiding framework for the entire project, elucidating the agreed-upon requirements that will shape the final product.

SPECIFICATION REQUIREMENT:

This segment depicts about both the practical and non useful necessity of the framework. The utilitarian prerequisite segment characterizes the framework outer interface, general necessity, execution, structure limitation and so forth.

GENERAL DESCRIPTIONS

The Software Requirements Specification (SRS) report holds a pivotal position, transcending the entirety of the project life cycle from its initiation to completion. Beyond being a static document, it evolves dynamically, adapting to the changing landscape of the software development process. This adaptability makes it a vital and enduring reference point, capturing the essence of the project's vision and requirements.

Throughout the project's progression, the SRS undergoes continuous updates and refinements, aligning with the evolution of the software's features and functionalities.

PRODUCT FUNCTION

The venture is ensured to give dependable outcomes and the usefulness of the item to recognize the misrepresentation exchanges viably and give adaptability to the client in a made sure about and precise way.

GENERAL CONSTRAINTS

- **Hardware Limitations:** There is no equipment limitation.
- **Interfaces to different Applications:** There will be no interfaces.
- **Audit Functions:** There will be no review capacities.
- **Control Functions:** There will be no control capacities

PRACTICAL ASPECTS

The correlation flanked by the information with yield to the framework is controlled by the practical necessity of the SRS(Software Requirement Specification).

TECHNICAL ISSUES

Numerous a product venture has flopped because of a deficient or mistaken examination process,particularly specialized issues. Specialized issues are a key advance while building up a product

RISK ANALYSIS

Task Risk Analysis is essential for precise cost assessments and risk evaluation on capital investment ventures. The primary challenge lies in determining the optimal way to represent and visualize the intricate interdependencies among risks, defining and monitoring the impactsof these risks, analyzing the likelihood of risk occurrence, mitigating the adverse effects of risks, and closely monitoring the project's progress in the presence of risks and vulnerabilities.

INTERFACE REQUIREMENTS

The framework execution is satisfactory. Be that as it may, Virtual travel organization is working with the client web association, 60% of the presentation is up to the customer side.

HARDWARE REQUIREMENTS

- Processor type — Pentium III-compatible processor
- RAM: — 512 MB or more
- HARD DISK — 20GB or more
- Monitor — VGA or higher resolution 800x600 or higher resolution
- Pointing device — Microsoft Mouse or compatible pointing device
- CD-ROM — Actual requirements will vary based on system configurationand the applications and features chosen to install.

SOFTWARE REQUIREMENTS

- Application software Framework : Python
  - Operating System: Windows XP Professional or above.

PERFORMANCE REQUIREMENTS

- The venture has the accompanying exhibition prerequisites.
- The prime necessity is that no blunder condition makes a venture exit suddenly.
- Any blunder happened in any procedure should restore a reasonable mistake message.

- The reaction ought to be genuinely quick, the activity members ought not be befuddledanytime of time about activity that is going on.

- The framework execution is sufficient.

SECURITY

The endeavor give a safety to diverse variety of clients by strategies for confirmation point. The endorsement apparatus of the system resolve frustrate the bothersome undertakings to the member of staff serving at table.

RELIABILITY

The project guarantees robust outcomes for the entire user base, ensuring a system uptime of 95%.

The commitment to excellence is reflected in the stringent criterion that the number of defects should not surpass 10 per functionality.

Moreover, prior to the submission of the final release, a thorough testing of the schedule will be conducted in the event that defects exceed the stipulated limit of 10 per functionality. This meticulousapproach ensures not only high system reliability but also a commitment to continuous quality assurance throughout the project's lifecycle.

USABILITY

- Since GUI interface is utilized, it tends to be utilized by a client.

- Since the framework is put on for online clients any sort client can utilize the framework.

- The framework recognizes the misrepresentation plus news of the client.

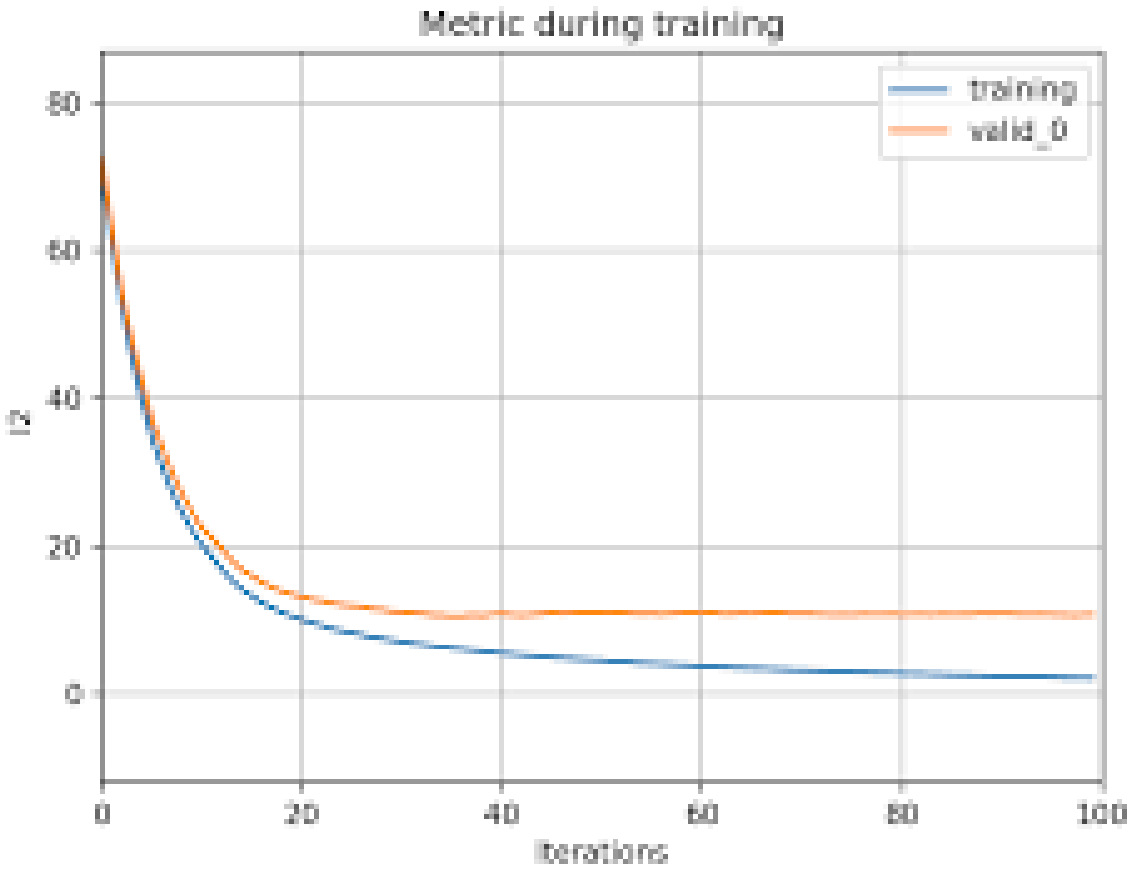
SCALABILITY

The prerequisite designed for flexibility have be a driver for an incredible section of the advancement improvements of the past very few existence. The trade has urban new encoding tongue, new arrangement frameworks, and original post and data move shows, somewhat to allow destinations toward create changing.

MAINTAINABILITY

Viability is our capacity to make changes to the item after some time. We need solid practicality so as to hold our initial clients. We will address this by envisioning a few kinds of progress, and via cautiously archiving our plan and usage.

ACURRACY GRAPH



CHAPTER 5

SYSTEM ANALYSIS AND SYSTEM DESIGN

This section provides comprehensive insights into the analysis conducted for the proposed system. System analysis serves as a crucial step in understanding the user requirements for the envisaged framework. It goes beyond merely capturing the current system's functionalities; it delves into the necessities that the proposed system aims to fulfill.

This includes a detailed exploration of the key features envisaged for the proposed framework and a thorough examination of the specific requirements that the system should address. By scrutinizing the existing system and identifying its limitations, this analysis lays the foundation for a robust and tailored solution that not only meets the current needs but also anticipates future requirements.

EXISTING SYSTEM

The conventional recognition techniques heavily lean on database systems and user training, often leading to delayed, inaccurate, and untimely results. Subsequently, methods grounded in distinctive analysis and regression analysis come into play, widely employed to identify fraudulent activities based on credit rating for applicants and credit card transactions. However, the efficiency of these methods diminishes significantly when handling large volumes of data.

In response to the limitations of traditional methods, there's a growing emphasis on leveraging advanced technologies such as machine learning, artificial intelligence, and data mining for credit card fraud detection.

PROBLEM RECOGNITION

The high measure of misfortunes because of extortion and the attention to the connection among misfortune and as far as possible must be diminished. The misrepresentation must be deducted continuously and the quantity of bogus alarm must be limited.

SUGGEST SYSTEM

The envisioned framework adeptly addresses the aforementioned challenge with efficiency and innovation. By leveraging the Light GBM algorithm, it not only identifies instances of fraud but also minimizes false alarms, yielding streamlined and precise outcomes. The crux of fraud detectionin this framework lies in scrutinizing user behavior, introducing a nuanced classification approach that incorporates a variable missorted cost.

The implementation of the Light GBM algorithm is a pivotal aspect of this framework, wherein numerous interim valued parameters are optimized. This optimization process ensures the algorithm's ability to discern patterns effectively, enhancing its capacity to distinguish fraudulent activities accurately. The adaptability and efficiency of the Light GBM algorithm make it an instrumental tool in mitigating fraud risks while delivering results with remarkable precision.

Furthermore, this framework aligns with the evolving landscape of fraud detection, acknowledging that user behavior is a dynamic variable. By integrating a comprehensive understanding of user behavior, the proposed system is poised to adapt to emerging fraud tactics and consistently provide reliable outcomes.

SYSTEM DESIGN

The process of configuration methodology involves the crucial steps of conceiving, planning, and structuring, with a primary emphasis on developing a conceptual framework, plan, or sketch. Design configuration serves as a vital stage where the conceptual representation of a proposed system transforms into tangible reality during development. Essential design factors, including reliability, response time, system throughput, efficiency, and scalability, must be carefully considered. Additionally, design objectives such as cost, hardware constraints, and compliance with standards should be thoroughly addressed.

In essence, the task of design configuration is to take the conceptual representation and translate it into a precise arrangement of facilities, personnel, equipment, and infrastructure to provide a comprehensive blueprint for a functional system. This entails the meticulous consideration of various elements, ensuring that the designed system aligns with the defined objectives and fulfills the desiredfunctionalities effectively.

ENGINEERED DESIGN

Illustrating the holistic attributes of the product places emphasis on delineating requirements and establishing the high-level architecture. In the subsequent design phase, various web pages and their interconnections are identified and meticulously planned.

The pivotal encoding components are identified and broken down into processing modules, logical data structures, andthe interconnected relationships among these modules are delineated.

[3,4,5]

FLOW OF LIGHTGBM ALGORITHM

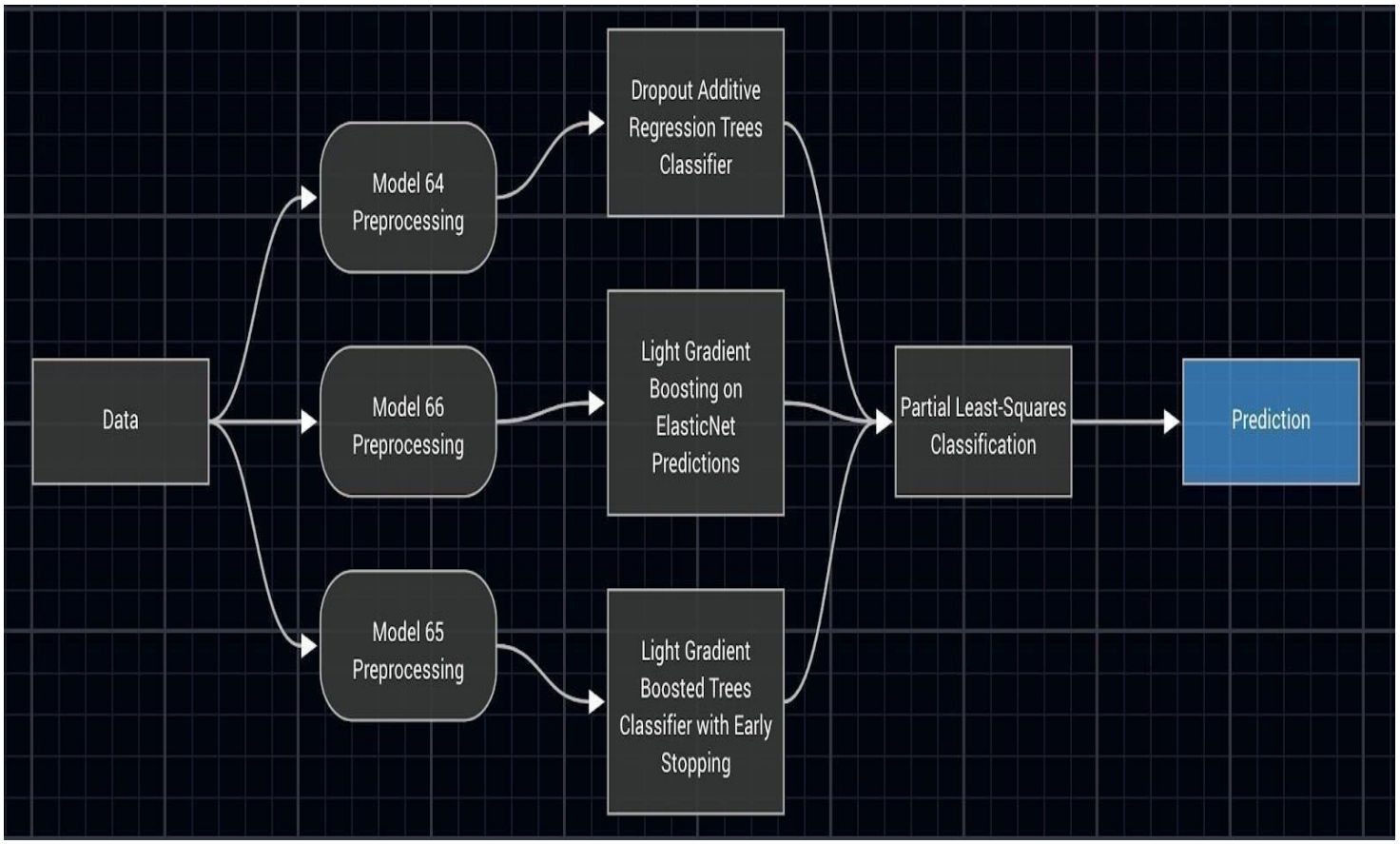


Figure 4. 3 FLOW DIAGRAMS

The above chart expresses the procedure of LightGbm calculation:

- Initially information is chosen arbitrarily from the datasets for preprocessing.
- In determination process is utilized by include Engineer to make relations.
- Then utilizes the model lightgbm for train the information.
- Then utilizes disarray lattice for information for make predications.

- Lightgbm the best arrangement are passed to the further age.

[3,4]

CHAPTER 6

CODING, TESTING AND USAGE

This is the way toward taking created arrangement of overhauled framework into operational use. On the off chance that the execution stage isn't painstakingly arranged and controlled, it can prompt numerous issues. Along these lines appropriate usage is basic to present an dependable framework to get together administrative supplies.

IMPLEMENTATION

This method was completely composed python. This empowers the charge card guarantors to utilize this method across wide assortment of gadgets autonomous of the merchant of the gadgets. We use prophet as a reverse finish for putting away dataset.

CODING

Standard coding rehearses are expected to guarantee that the code is intelligible, reasonable and effectively modifiable. This venture has characterized gauges and rules to be followed while pseudo coding. These principles were followed during the improvement of the application to deliver code that is progressively predictable and to create code.

TESTING

Trying is single stage in the item/net structure process that possibly will be viewed as destructives instead of important. Testing requires that the originator discard

Suspicions of the correctness of the item just made and beat a hopeless situation.

In addition data assembled as testing is driven offer a not too bad hint of programming trustworthiness and some indication of programming quality as total. Testing can't show the nonappearance of defects, it can simply show that item absconds that are accessible.

UNIT TESTING

Designers conduct unit tests to validate their individual components against established guidelines.Unit testing stands apart from integration testing, which verifies that components function correctly together, and acceptance testing, which ensures that a feature performs as expected by the client. The term "unit" in unit tests signifies that they examine a singular unit of code. This testing approach places emphasis on confirming the correctness of the smallest coding units.

Unit testing is aptly named because it focuses on scrutinizing individual components or modules of the software. This testing process directs verification efforts towards the smallest building blocks of the coding structure. Each module within this project undergoes independent scrutiny to identify and rectify any errors present.

CHAPTER 7

CONCLUSION AND FORESEEABLE ENHANCEMENTS

CONCLUSION

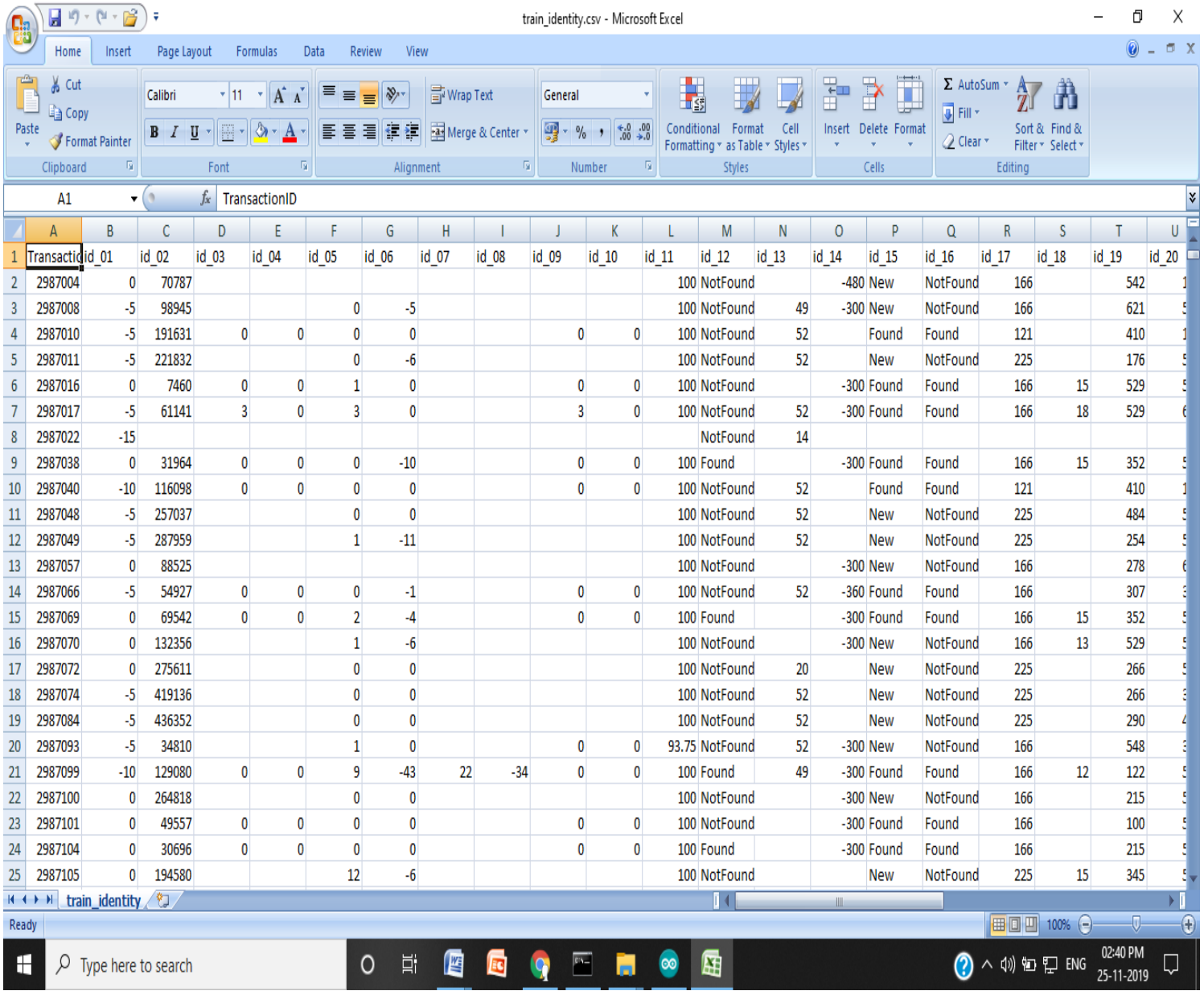
This approach proves to be highly effective in accurately detecting fraudulent transactions while simultaneously minimizing the occurrence of false alarms. The utilization of the LightGBM algorithm introduces a novel dimension to the realm of fraud detection, particularlywithin the application space discussed. When applied within the credit card fraud detection systems of banking institutions, the LightGBM algorithm exhibits a remarkable capability to predict instances of fraudulent transactions well in advance of standard credit card transactions.

The integration of this algorithm opens up opportunities to implement a series of anti-fraud strategies within the banking sector. By anticipating potential fraudulent transactions with higher precision, financial institutions can proactively adopt preventive measures to mitigate the risk of substantial losses. These anti-fraud strategies not only serve to safeguard the interests of banks but also contribute to the overall reduction of risks associated with fraudulent activities.

FUTURE ENHANCEMENTS

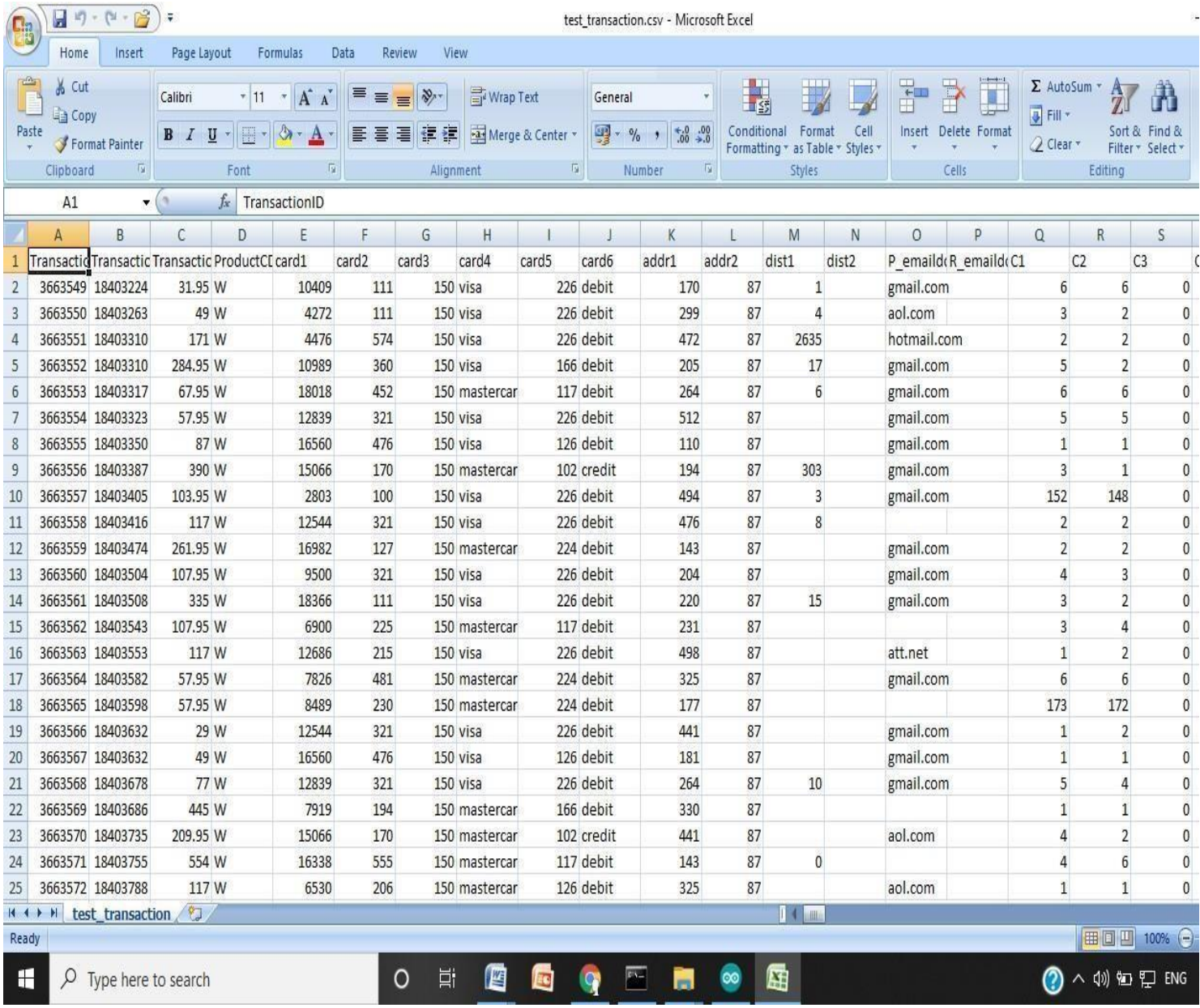
The discoveries got here will not add to the worldwide extortion identification problem. As future did some successful work calculation which can execute fine meant for the characterization subject by means of unpredictable misclassification cost possibly will be created

APPENDIX SCREEN SHOTS



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Transactid	id_01	id_02	id_03	id_04	id_05	id_06	id_07	id_08	id_09	id_10	id_11	id_12	id_13	id_14	id_15	id_16	id_17	id_18	id_19	id_20
2	2987004	0	70787									100	NotFound		-480	New	NotFound	166		542	1
3	2987008	-5	98945			0	-5					100	NotFound	49	-300	New	NotFound	166		621	5
4	2987010	-5	191631	0	0	0	0			0	0	100	NotFound	52		Found	Found	121		410	1
5	2987011	-5	221832			0	-6					100	NotFound	52		New	NotFound	225		176	5
6	2987016	0	7460	0	0	1	0			0	0	100	NotFound		-300	Found	Found	166	15	529	5
7	2987017	-5	61141	3	0	3	0			3	0	100	NotFound	52	-300	Found	Found	166	18	529	6
8	2987022	-15											NotFound	14							
9	2987038	0	31964	0	0	0	-10			0	0	100	Found		-300	Found	Found	166	15	352	5
10	2987040	-10	116098	0	0	0	0			0	0	100	NotFound	52		Found	Found	121		410	1
11	2987048	-5	257037			0	0					100	NotFound	52		New	NotFound	225		484	5
12	2987049	-5	287959			1	-11					100	NotFound	52		New	NotFound	225		254	5
13	2987057	0	88525									100	NotFound		-300	New	NotFound	166		278	6
14	2987066	-5	54927	0	0	0	-1			0	0	100	NotFound	52	-360	Found	Found	166		307	3
15	2987069	0	69542	0	0	2	-4			0	0	100	Found		-300	Found	Found	166	15	352	5
16	2987070	0	132356			1	-6					100	NotFound		-300	New	NotFound	166	13	529	5
17	2987072	0	275611			0	0					100	NotFound	20		New	NotFound	225		266	5
18	2987074	-5	419136			0	0					100	NotFound	52		New	NotFound	225		266	3
19	2987084	-5	436352			0	0					100	NotFound	52		New	NotFound	225		290	4
20	2987093	-5	34810			1	0			0	0	93.75	NotFound	52	-300	New	NotFound	166		548	3
21	2987099	-10	129080	0	0	9	-43	22	-34	0	0	100	Found	49	-300	Found	Found	166	12	122	5
22	2987100	0	264818			0	0					100	NotFound		-300	New	NotFound	166		215	5
23	2987101	0	49557	0	0	0	0			0	0	100	NotFound		-300	Found	Found	166		100	5
24	2987104	0	30696	0	0	0	0			0	0	100	Found		-300	Found	Found	166		215	5
25	2987105	0	194580			12	-6					100	NotFound			New	NotFound	225	15	345	5

Figure 7. 1 Dataset 1



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Transactio	Transactio	Transactio	ProductCI	card1	card2	card3	card4	card5	card6	addr1	addr2	dist1	dist2	P_emaildo	R_emaildo	C1	C2	C3
2	3663549	18403224	31.95	W	10409	111	150	visa	226	debit	170	87	1		gmail.com		6	6	0
3	3663550	18403263	49	W	4272	111	150	visa	226	debit	299	87	4		aol.com		3	2	0
4	3663551	18403310	171	W	4476	574	150	visa	226	debit	472	87	2635		hotmail.com		2	2	0
5	3663552	18403310	284.95	W	10989	360	150	visa	166	debit	205	87	17		gmail.com		5	2	0
6	3663553	18403317	67.95	W	18018	452	150	mastercar	117	debit	264	87	6		gmail.com		6	6	0
7	3663554	18403323	57.95	W	12839	321	150	visa	226	debit	512	87			gmail.com		5	5	0
8	3663555	18403350	87	W	16560	476	150	visa	126	debit	110	87			gmail.com		1	1	0
9	3663556	18403387	390	W	15066	170	150	mastercar	102	credit	194	87	303		gmail.com		3	1	0
10	3663557	18403405	103.95	W	2803	100	150	visa	226	debit	494	87	3		gmail.com		152	148	0
11	3663558	18403416	117	W	12544	321	150	visa	226	debit	476	87	8				2	2	0
12	3663559	18403474	261.95	W	16982	127	150	mastercar	224	debit	143	87			gmail.com		2	2	0
13	3663560	18403504	107.95	W	9500	321	150	visa	226	debit	204	87			gmail.com		4	3	0
14	3663561	18403508	335	W	18366	111	150	visa	226	debit	220	87	15		gmail.com		3	2	0
15	3663562	18403543	107.95	W	6900	225	150	mastercar	117	debit	231	87					3	4	0
16	3663563	18403553	117	W	12686	215	150	visa	226	debit	498	87			att.net		1	2	0
17	3663564	18403582	57.95	W	7826	481	150	mastercar	224	debit	325	87			gmail.com		6	6	0
18	3663565	18403598	57.95	W	8489	230	150	mastercar	224	debit	177	87					173	172	0
19	3663566	18403632	29	W	12544	321	150	visa	226	debit	441	87			gmail.com		1	2	0
20	3663567	18403632	49	W	16560	476	150	visa	126	debit	181	87			gmail.com		1	1	0
21	3663568	18403678	77	W	12839	321	150	visa	226	debit	264	87	10		gmail.com		5	4	0
22	3663569	18403686	445	W	7919	194	150	mastercar	166	debit	330	87					1	1	0
23	3663570	18403735	209.95	W	15066	170	150	mastercar	102	credit	441	87			aol.com		4	2	0
24	3663571	18403755	554	W	16338	555	150	mastercar	117	debit	143	87	0				4	6	0
25	3663572	18403788	117	W	6530	206	150	mastercar	126	debit	325	87			aol.com		1	1	0

Figure 7. 2 Dataset 2

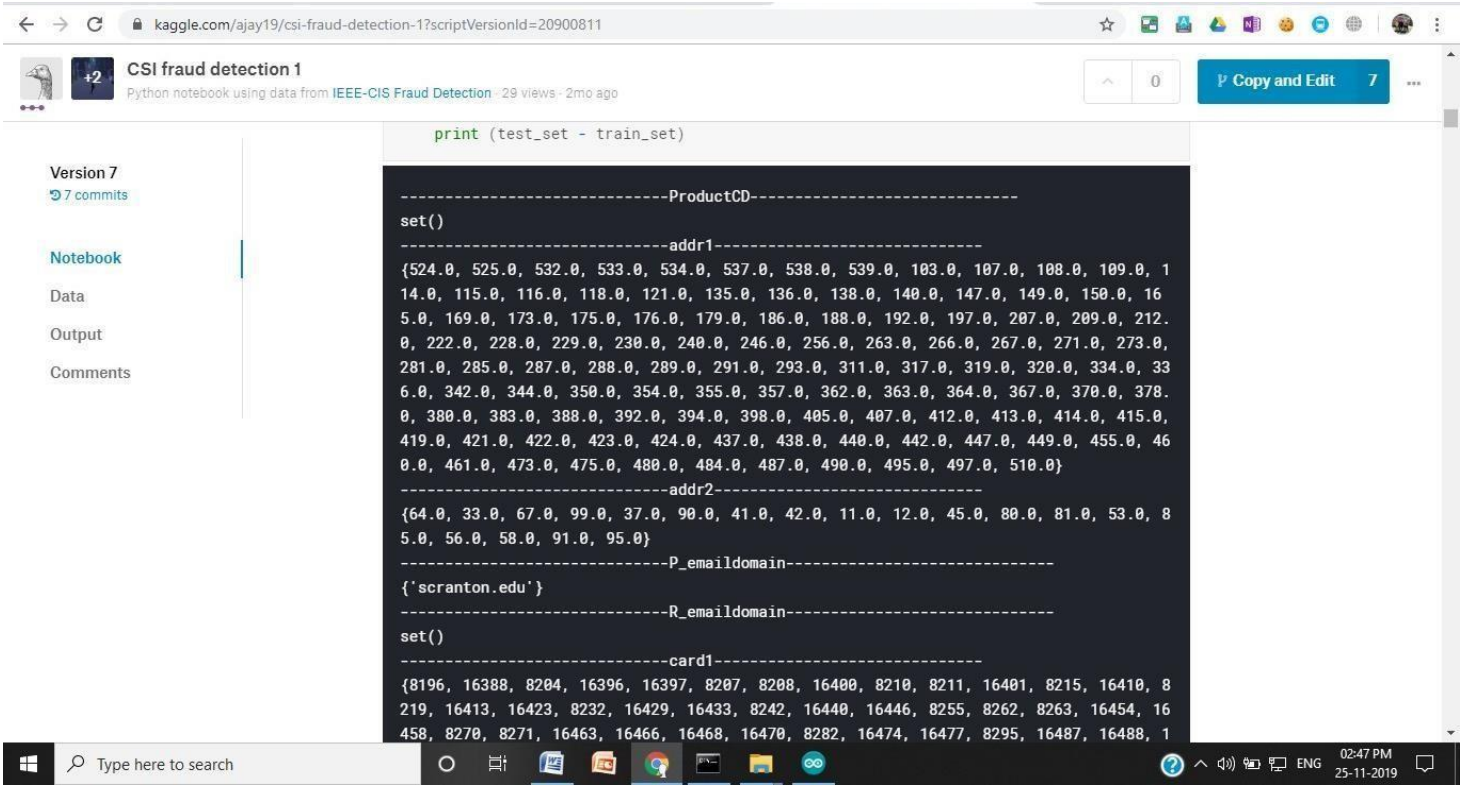


Figure 7. 3 Filtering the Data

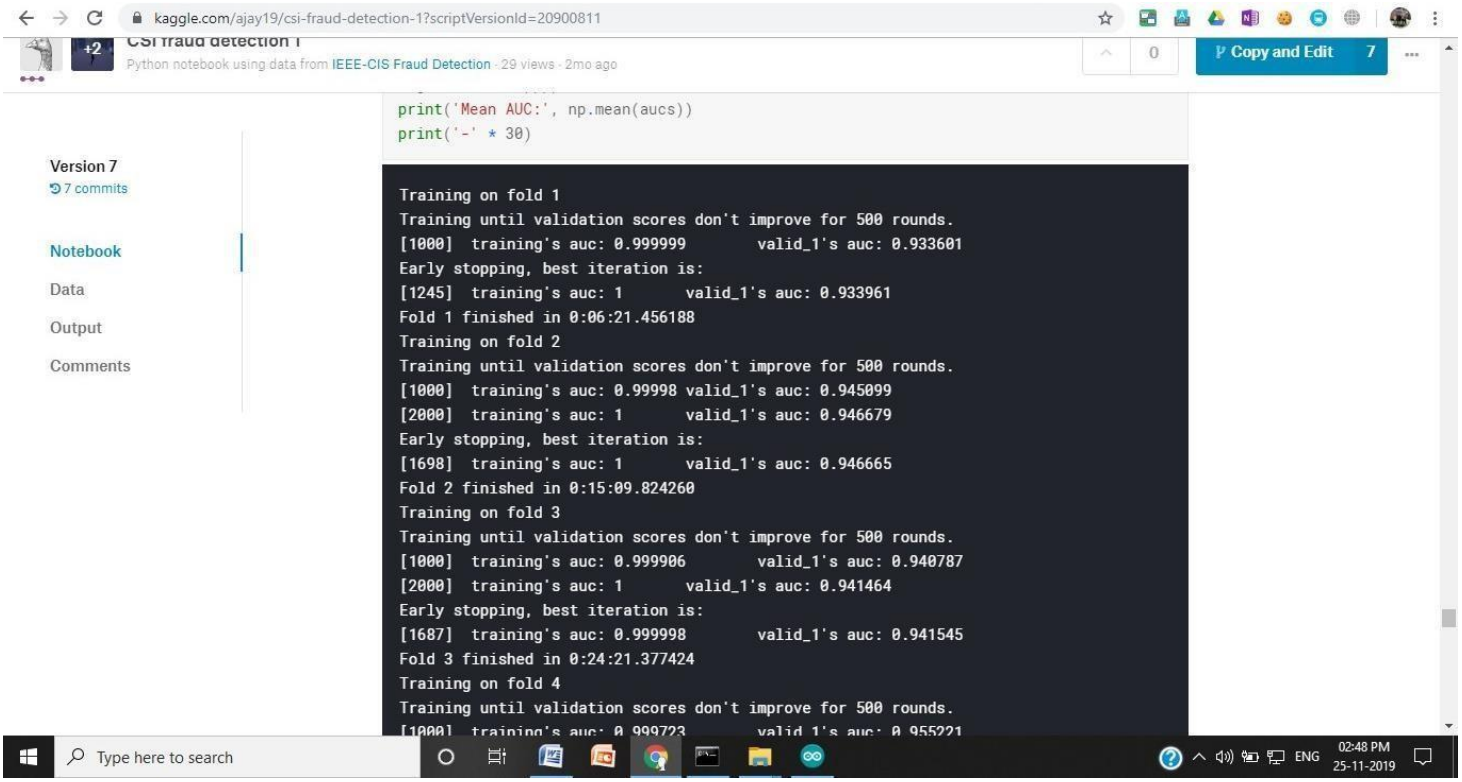


Figure 7. 4 Train the Model

BIBLIOGRAPHY
REFERENCES

Papers:

[1] M. Hamdi Ozcelik, Ekrem Duman, Mine Isik, Tugba Cevik, Improving a credit card fraud detection system using Lighgbm algorithm, International conference on Networkingand information technology 2010.

[2] Wen-Fang YU, Na Wang, Research on Credit Card Fraud Detection Model Based on Distance Sum, IEEE International Joint Conference on Artificial Intelligence 2009.

[3] clifton phua, vincent lee1, kate smith & ross gayler, A Comprehensive Survey of Data Mining-based Fraud Detection Research,2005.

[4] Elio Lozano, Edgar Acu˜na, Parallel algorithms for distance-based and density-based outliers,2006.

[5] Credit card fraud detection using hidden markov model – Abinav Srivastava,Amlan Kundu,Shamik Sural,Arun K.majumdar

Websites:

[1] http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/tcw2/report.html

[2] http://www.kxcad.net/cae_MATLAB/toolbox/gads/f6691.html

[3] http://www.easywayserver.com/blog/user-login-in-jsp/

[4]http://www.faqs.org/patents/app/20100094765

Textbooks:

[1]Pressman, Roger S. Software engineering: a practitioner‘s approach / Roger S. Pressman.—5th ed.p. cm (McGraw-Hill series in computer science).

[2]E.Balagurasamy, Programming with python, Tata McGraw-Hill Publication.

[3]Ali Bahrami,Oject Oriented system Development, Tata McGraw-Hill Publication . [4]Jiwei Han et,al., –Data Mining :Concepts and Techniques‖, MorganKaufmaan Series,2000