Report On

Credit Card Fraud Detection

Submitted in partial fulfillment of the requirements of the Mini project in
Semester VII of Final Year Artificial Intelligence And Data Science

by
Ujwal Sharma (26)
Om Tiwari (29)
Riddhesh Sankhe (36)

Guide
Prof. Bhavika Gharat

University of Mumbai

Vidyavardhini's College of Engineering & Technology

Department of Artificial Intelligence And Data Science

(A.Y. 2023-24)

Vidyavardhini's College of Engineering & Technology

Department of Artificial Intelligence And Data Science

CERTIFICATE

This is to certify that the Mini Project entitled **"Real Time Object Detection"** is a bonafide work of **Ujwal Sharma (26), Om Tiwari (29)** and **Riddhesh Sankhe (36)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **"Bachelor of Engineering"** in Semester VII of Final Year **"Artifical Intelligence And Data Science"**.

Supervisor

Prof. Bhavika Gharat

Dr. Tatwadasrhi P.N.
Head of Department

Contents

Abstract

We are implementing a credit card fraud detection system, by using big data technologies. Credit card is one of the most divisive products among all the financial tools available. The usage of credit cards has become common in today's world and huge volume of transaction happens online. The increase in these transactions has also come with many apprehensions on the authenticity of the transactions. In today's world, there have been various phishing attacks over the internet. This needs to be dealt with caution. Be focus on designing an online credit card fraud detection framework with Big data technology. To accomplish that, we propose a workflow which satisfies most design ideas of current credit card fraud detection systems. We implement it with largest Big data technologies like Hadoop, Spark, Apache Kafka etc. A prototype is implemented and tested with a synthetic dataset, which shows great potentials of achieving the above goals.

Acknowledgements

1. Introduction

1.1. Introduction

Data is being increased at an exponential rate every second with the rapid development of e-commerce and internet. The banking sector has become a very important sector in our present generation where every human has to deal with bank either online or physically. With this development, online transactions has become one of the most important ways of trading. Credit card is one of the ways used for performing online transaction. A credit card is nothing but a payment card issued to users to enable the cardholder to pay a merchant for the required goods and services. In dealing with banks and credit cards, the customers face the chances of being trapped by fraudsters. In online transaction usually the credit cards are used as virtual cards. An fraudster only needs to know a few important details of the card like card id, security code etc to make the fraudulent transaction. While the genuine card holder often does not realise that their credit card details has been leaked, which causes a significant financial loss to both the cardholder and the credit card company.In the decades, many researchers have developed many credit card fraud detection systems. The main challenge in most of the ccfds is improving the detection accuracy and the capacity of computation with the explosive growth of trading data. As the number of users are being increasing day by day , the transactions has bought heavy workloads to these detection systems. The number of transactions per second can reach upto millions and the historic data stored can reach upto PBs or even EBs.

1.2. Problem statement & Objectives

Problem Statement: Develop an advanced Credit Card Fraud Detection system leveraging Big Data technologies to accurately and efficiently detect fraudulent transactions in real-time. The system should analyze large volumes of transaction data to identify patterns and anomalies, reducing false positives while maximizing the detection of genuine fraud attempts. The solution should be scalable, cost-effective, and capable of handling the evolving tactics of fraudsters in a rapidly changing financial landscape.

Objectives:

- Real-time Detection: Develop a system that can identify fraudulent transactions as they occur, minimizing the financial impact on cardholders and financial institutions.

- High Accuracy: Achieve a high level of accuracy in fraud detection to minimize false positives, ensuring that genuine transactions are not declined.
- Scalability: Build a system that can scale horizontally to handle the ever-increasing volume of credit card transactions and adapt to future growth.
- Anomaly Detection: Utilize big data analytics to identify anomalies and patterns in transaction data that may indicate fraud, even in cases with no known fraud history.
- Cost Efficiency: Design the solution to be cost-effective, optimizing resource utilization and minimizing

1.3. Scope:

The scope of the Credit Card Fraud Detection using Big Data project encompasses the development of a robust, real-time fraud detection system that leverages large volumes of credit card transaction data. This project will involve data collection, preprocessing, and integration, implementing machine learning models for fraud detection, as well as predictive analytics and adaptive learning to identify emerging fraud patterns. It includes the use of Big Data technologies for efficient data processing, alert mechanisms for notifying relevant stakeholders, model evaluation, compliance with data protection laws, scalability, integration with existing financial infrastructure, incident response procedures, customer communication strategies, training, and security measures. The project aims to create a comprehensive, adaptive, and compliant solution for detecting and preventing credit card fraud while continually improving its performance and adhering to legal and ethical considerations.

2. Literature Survey

A literature survey for a credit card fraud detection system report would involve researching and summarizing relevant academic papers, articles, and publications on the topic. Here's a brief outline of how you can approach this Start with an introduction that explains the importance of credit card fraud detection and its impact on financial institutions and consumers.Describe the search strategy you used to identify relevant literature, including keywords, databases, and any specific criteria for inclusion. Provide a brief historical overview of credit card fraud detection systems to show how they have evolved over time. Summarize research on machine learning and data mining techniques used in credit card fraud detection. Mention popular algorithms like Random Forest, Neural Discuss methods for feature selection and engineering in fraud detection systems to highlight relevant variables. Address the issue of imbalanced data and methods used to tackle it, such as oversampling, undersampling, and cost-sensitive.

3. Methodology

3.1. Introduction:

 Fraud detection is a critical concern for financial institutions, e-commerce platforms, and various industries that process vast amounts of transactional data. In an increasingly interconnected and digital world, the need for robust fraud detection systems has never been more pressing. Big Data Analytics (BDA) has emerged as a pivotal technology in this realm, offering the capabilities to process and analyze enormous volumes of data swiftly and effectively.This methodology aims to outline the approach and strategies for building a cutting-edge fraud detection system empowered by Big Data Analytics. Traditional fraud detection systems are often limited by their ability to handle substantial data loads, leading to the omission of valuable insights and potentially dangerous false negatives. In contrast, BDA equips organizations with the tools necessary to harness the full potential of their data.This methodology will guide the design, implementation, and deployment of a robust fraud detection system that leverages BDA to address these challenges. By harnessing the power of BDA, we can identify and thwart fraudulent activities in real-time while minimizing false positives, thus safeguarding financial resources, data integrity, and user trust.

3.2. Architecture/ Framework/ Block diagram:
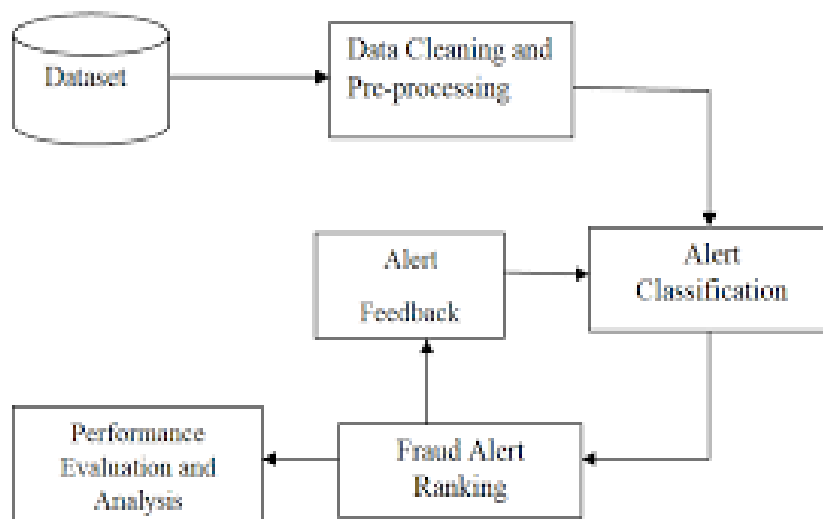

Fig 1: Process Design

4

3.4. Details of hardware & software:

Hardware Requirements:

- I3 Processor Based Computer or higher

- Memory: 2 GB RAM

- Hard Drive: 50 GB

- Monitor

- Internet Connection

Software Requirements:

- Windows 7 or higher

- Programming language: Python

- IDE used: Jupyter notebook

4. Result
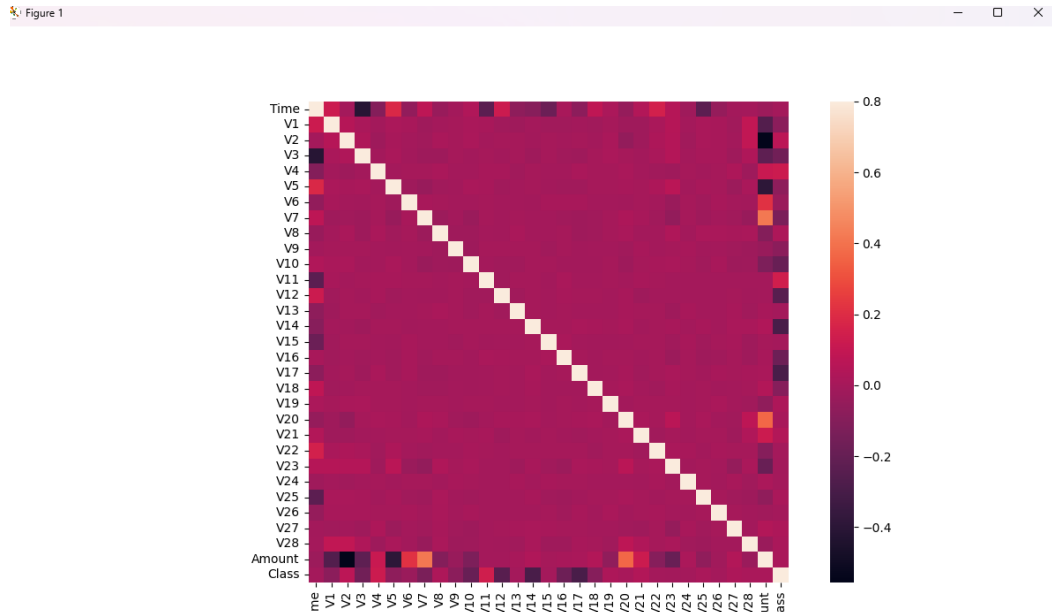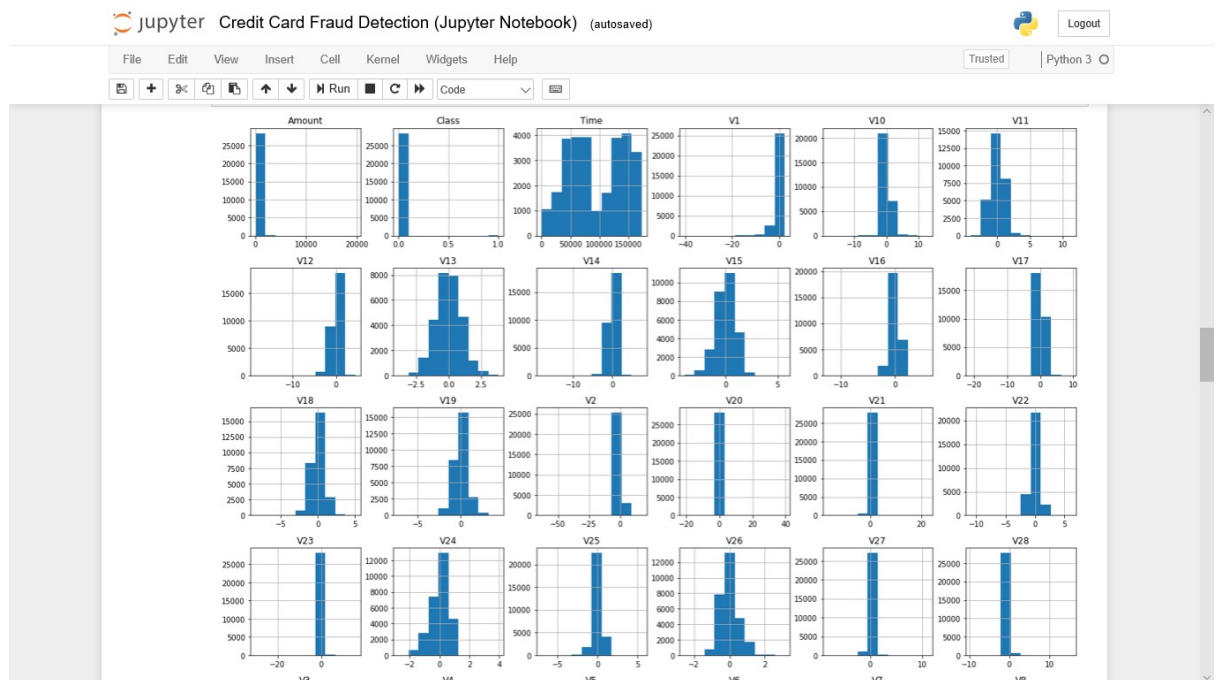
4.1. Output

Fig 1. Confusion Matrix

```python
# V1 - V28 are the results of a PCA Dimensionality reduction to protect user identities and sensitive features
```

```
(28481, 31)
              Time            V1            V2            V3            V4  \
count  28481.000000  28481.000000  28481.000000  28481.000000  28481.000000
mean   94705.035216     -0.001143     -0.018290      0.000795      0.000350
std    47584.727034      1.994661      1.709050      1.522313      1.420003
min        0.000000    -40.470142    -63.344698    -31.813586     -5.266509
25%    53924.000000     -0.908809     -0.610322     -0.892884     -0.847370
50%    84551.000000      0.031139      0.051775      0.178943     -0.017692
75%   139392.000000      1.320048      0.792685      1.035197      0.737312
max   172784.000000      2.411499     17.418649      4.069865     16.715537

                 V5            V6            V7            V8            V9  \
count  28481.000000  28481.000000  28481.000000  28481.000000  28481.000000
mean      -0.015666      0.003634     -0.008523     -0.003040      0.014536
std        1.395552      1.334985      1.237249      1.204102      1.098006
min      -42.147898    -19.996349    -22.291962    -33.785407     -8.739670
25%       -0.703986     -0.765807     -0.562033     -0.208445     -0.632488
50%       -0.068037     -0.269071      0.028378      0.024696     -0.037100
75%        0.603574      0.398839      0.559428      0.326057      0.621093
max       28.762671     22.529298     36.677268     19.587773      8.141560

       ...           V21           V22           V23           V24  \
count  ...  28481.000000  28481.000000  28481.000000  28481.000000
mean   ...      0.004740      0.006719     -0.000494     -0.002626
std    ...      0.744743      0.728209      0.645945      0.603968
min    ...    -16.640785    -10.933144    -30.269720     -2.752263
25%    ...     -0.224842     -0.535877     -0.163047     -0.360582
50%    ...     -0.029075      0.014337     -0.012678      0.038383
75%    ...      0.189068      0.533936      0.148065      0.434851
max    ...     22.588989      6.090514     15.626067      3.944520

                V25           V26           V27           V28        Amount  \
count  28481.000000  28481.000000  28481.000000  28481.000000  28481.000000
mean      -0.000917      0.004762     -0.001689     -0.004154     89.957884
std        0.520679      0.488171      0.418304      0.321646    270.894630
min       -7.025783     -2.534330     -8.260909     -9.617915      0.000000
25%       -0.319611     -0.328476     -0.071712     -0.053379      5.980000
```

5. Conclusion

An exceedingly proficient and exact credit card fraud detection framework is the need of great importance as a many transactions are being done each day. Thus a expansive measure of research is being done in this area furthermore, various systems are proposed conquered credit card extortion. The Fraud Detection System is additionally versatile for taking care of tremendous volumes of transactions. The HMM based credit card fraud detection framework isn't taking long time what's more, having complex procedure to perform misrepresentation check like the existing framework and it gives preferable and quick come about over existing framework. The Hidden Markov Model makes the handling of recognition simple and tries to expel the many-sided quality. In this paper, big data technologies like Spark, Kafka and zookeeper are being used for storing and processing of large amount of historical transactions of the users. Our result shows the effectiveness and correctness of the proposed system over a broad deviation of data.

6. References

[1] ACI'21: Workshop on Advances in Computational Intelligence at ISIC 2021,February 25–27, 2021, New Delhi, India

[2]Sandeep Kumar, Aman Balyan, Manvi Chawla Computer Science &Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi, India. © 2017 IJEDR | Volume 5, Issue 4 | ISSN: 2321-9939

[3]Saranya, N., Nandinipriya, M., & Priya, U. (2018). Real time object detection for blind people. International Journal of Advance Research in Science and Engineering, 7(1), 306-316

4]] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).