B. P. Poddar Institute of Management and Technology

Department of Computer Science & Engineering

Project Synopsis

Title: Credit Card Fraud Detection System

Group No - 8:

Name	Roll
Sagar Debnath	11500120009
Ramesh Das	11500120010
Shoham Sen	11500120075
Soudeep Ghosh	11500120093

Title: Credit Card Fraud Detection System

Abstract/ Project definition:

In recent years, the advancements of e-commerce and e-payment systems have resulted in a rise in financial fraud cases, such as credit card fraud. It is therefore crucial to implement mechanisms that can detect credit card fraud. Here, comes the need for a system that can track the pattern of all the transactions and if any pattern is abnormal then the transaction should be aborted.Today, we have many machine learning algorithms that can help us classify abnormal transactions.But the features of credit card frauds play an important role when machine learning is used for credit card fraud detection, and they must be chosen properly.To validate the performance, we will use a dataset that contains transactions made by credit cards in September 2013 by European cardholders.

Literature review:

Gupta and his team developed an automated model to detect economically related fraudulent instances, with a focus on credit card transactions. Among the various ML techniques used, Naïve Bayes performed exceptionally well, with an accuracy of 80.4% and an area under the curve of 96.3% (Gupta et al., 2021). [1]

Mailini and Pushpa proposed using KNN and anomaly detection to detect credit card fraud, and the authors, after completing the sample data, found that the method was KNN as the best way to detect and identify flaws in Target. the best. to identify fake memory. Credit card verification requires less computation and memory for suspicious detection and works faster and better on large online databases. However, his studies and results show that KNN is accurate and effective (Malini & Pushpa, 2017). [2]

Varmedja's team uses various machine learning algorithms in their paper, such as logistic regression, multilayer perceptron, random forest and pure Bayesian. Because the data was inconsistent, Varmedja and his team used SMOTE techniques for oversampling, feature selection, and further partitioning of data into training and test datasets. The model with the best score during the test is Random Forest with a score of 99.96%, not much different, with Multilayer Perceptron in second place with a score of 99.93% and Naive Bayes in third place with a score of 99.23% according to Logistic Regression with 97.46% (Varmedja et al., 2019). [3]

Kiran and his team briefly present the K-Neighbor Neighbor credit card fraud detection method (NBKNN) enhanced with Naive Bayesian (NB). Experimental results show the difference in performance of each classifier in the same dataset. Naive Bayes outperforms K Neighbors as it is 95% accurate compared to 90% for KNN (Kiran et al., 2018). [4]

Itoo and his team's work used three different machine learning methods, the first is logistic regression, the second is Naive Bayes, and the last is the K-Best approximation. Itoo and his team documented and compared their work with python. Logistic Regression has an accuracy of 91.2%, Naive Bayes has an accuracy of 85.4%, and the K-Nearest is the closest with an accuracy of 66.9% (Ito et al., 2020). [5]

Expected outcome:

The expected outcome of a credit card fraud detection system using machine learning (ML) techniques is an improved ability to accurately detect credit card fraud while minimizing false positives. Trained on historical data that includes both legitimate and fraudulent transactions, ML algorithms can learn patterns and anomalies that may indicate fraudulent activity. The expected results of the system are as follows:

1. Improved detection accuracy: ML algorithms can quickly process large amounts of data and detect subtle patterns that traditional rule-based systems may miss. This should lead to more accurate detection of fraudulent transactions.
2. Fewer false positives: Traditional rule-based systems often generate false positive alerts and flag legitimate transactions as potentially fraudulent. A well-trained ML system should be able to distinguish genuine from suspicious events more efficiently and reduce the number of false alarms.

3. Adaptability: ML-based systems can adapt to new types of fraud as criminals develop new tactics. The system should be able to learn and evolve to detect new fraud patterns.
4. Detection in real time: ML algorithms can process transactions in real time, enabling rapid detection and response to potential fraud, minimizing the financial impact on both cardholders and card issuers.
5. Continuous Improvement: ML models can continuously learn from new data and improve their ability to distinguish between legitimate and fraudulent transactions over time.
6. Scalability: ML-based fraud detection systems can be scaled to handle increasing transaction volumes without compromising detection accuracy.

It is important to note that while ML-based systems offer significant advantages, they are not foolproof and must be constantly monitored and updated to adapt to evolving fraud techniques. Collaboration between industry experts, data scientists and fraud analysts is critical to the continued effectiveness of the system.

Timeline:

Task	16th-31st Jul	1𝑠𝑡-15𝑡ℎ Aug	16th-31st Aug	1st-30th Sep	1st-31st Oct	1st-10th Nov	11th-30th Nov	1st-20th Dec
Project definition	↔							
Literature Review		↔						
Data Collection		↔						
Familiarization with Machine learning		↔	↔	↔				
Familiarization with Python				↔	↔			
Data cleaning and preprocessing					↔			
Model Development and Training					↔	↔		
Feature Selection and Optimization							↔	
System Integration and Testing								↔
Report writing and project presentation								↔

Program Outcomes (POs):

1. Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. Problem analysis: Identify, formulate, research literature, and analyses complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. The engineer and society: Apply to reason informed by the contextual knowledge to health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of and need for sustainable development.
8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. Individual and teamwork: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work,

as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PO/PSO mapping:

PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO 10	PO 11	PO 12	PSO 1	PSO 12
3	3	3	2	3	2	1	2	3	3	-	1	3	3

References:

[1] Gupta, A., Lohani, M. C., & Manchanda, M. (2021). Financial fraud detection using naive Bayes algorithm in highly imbalance data set. Journal of Discrete Mathematical Sciences and Cryptography, 24(5), 1559–1572. https://doi.org/10.1080/09720529.2021.1969733

[2] Malini, N., & Pushpa, M. (2017). Analysis on credit card fraud identification techniques based on KNN and outlier detection. 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB). https://doi.org/10.1109/aeeicb.2017.7972424

[3] Varmedja, D., Karanovic, M., Sladojevic, S., Arsenovic, M., & Anderla, A. (2019). Credit Card Fraud Detection - machine learning methods. 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH). https://doi.org/10.1109/infoteh.2019.8717766

[4] Kiran, S., Guru, J., Kumar, R., Kumar, N., Katariya, D., & Sharma, M. (2018). Credit card fraud detection using Naïve Bayes model based and KNN classifier.

International Journal Of Advance Research, Ideas And Innovations In Technology, 4(3).

[5] Itoo, F., Meenakshi, & Singh, S. (2020). Comparison and analysis of logistic regression, Naïve Bayes and Knn Machine Learning Algorithms for credit card fraud detection. International Journal of Information Technology, 13(4), 1503–1511. https://doi.org/10.1007/s41870-020-00430-y