CS 5800: Algorithms Final Project Proposal

Credit Card Fraud and Anomaly Detection Using Clustering Algorithms

Abhinav Nippani, Adwait Patil, Ayush Patel, Shivansh Verma

Problem Context in view of Abhinav Nippani

In the dynamic landscape of digital transactions, the increase in online financial activities has exposed individuals to the looming threat of credit card fraud, endangering personal financial security and the stability of the financial ecosystem. Driven by a commitment to contribute to technological solutions that address real-world challenges, my interest in the project springs from a deep-rooted passion to harness innovative tools for safeguarding financial transactions, based on my previous work experience as a business analyst in finance domain. The implementation of clustering algorithms such as K-Means, KNN, and DBSCAN from scratch is particularly engaging, providing a practical application of the knowledge acquired during my coursework, aiming not only to enhance fraud detection but also to deepen my understanding of financial data analytics. This project aligns with my academic aspirations and serves as a significant step toward contributing valuable insights to the ongoing battle against credit card fraud in the Indian context. It is a pivotal endeavour in my personal and professional journey, offering an opportunity to make a positive impact at the intersection of technology and finance in the Indian scenario.

Problem Context in view of Adwait Patil

The surge in cybercrimes accompanying the global shift to digital financial transactions mandates a proactive response to mitigate risks, with credit card fraud standing out as a pressing concern. This project addresses the imperative need for an advanced Credit Card Fraud Detection system using clustering algorithms—specifically, KNN, K-Means, and DBSCAN. By employing KNN, the project aims to discern transaction patterns by analysing the proximity of instances in a feature space, aiding in anomaly detection. K-Means clustering facilitates the grouping of transactions based on similarities, enhancing the identification of fraudulent activities. Additionally, DBSCAN's capability to detect dense regions provides valuable insights into irregularities. A pivotal aspect of this project involves optimising the Expectation-Maximization (EM) step within the K-Means algorithm. The EM step's refinement of cluster assignments and centroid updates significantly influences overall algorithm performance. Integration of advanced optimization algorithms in this step seeks to enhance efficiency, contributing to a more accurate credit card fraud detection system.

Problem Context in view of Ayush Patel

In this digital age, the surge of online financial transactions has significantly increased the vulnerability to credit card fraud, posing a serious risk to both personal financial security and the stability of the financial ecosystem at large. Having personally faced the challenges and aftermath of credit card fraud, I've become acutely aware of the critical need for stronger protective measures in financial dealings. This direct encounter with fraud not only heightened my awareness of the existing gaps in financial security systems but also ignited my interest in leveraging technological solutions to combat such risks. Our project uses K-Means, KNN and DBSCAN which we will implement from scratch and will be used to detect fraud patterns in credit cards. I am keen on using DBSCAN which we can implement both from a BFS or DFS approach. By doing this project, I can apply these algorithms from scratch which I learned from this course to solve the real problem we are facing today, which are challenging and meaningful. I believe this project will be valuable to me.

Problem Context in view of Shivansh Verma

In our day-to-day life we often get caught up in a lot of fast paced action items that we often fail to check the most basic things that impact our everyday life. We often protect our savings and implement the strictest measures to avoid any cyber-attacks or any unwarranted losses. Oftentimes we are surprised by some transactions that we did not authorise and are surprised by it when we see it; but by the time we see some suspicious activity it is too late to raise the alarms and recoup the money lost. This time delay is the reason for most people losing their money, people often see fraudulent transactions at a stage where it is near impossible to find the culprits and file complaints that will result in some actionable results. Therefore, we need to find an optimum clustering solution that can get results in a faster and more accurate manner. I want to better understand the places where we can optimise the time and space complexity as clustering algorithms are notoriously difficult to crack. A simple KNN algorithm runs in n-squared time complexity which is very bad for a real-world problem. Therefore, I wish to use the concepts learnt during the course to optimise these algorithms.

Question:

For our project the central question we aim to answer is: How can clustering algorithms like DBSCAN, KMeans, and KNN be effectively implemented to detect and prevent fraudulent activities in credit card transactions? This inquiry delves into the practical application of these algorithms to accurately identify anomalous transaction patterns that deviate from normal behaviour, thus signalling potential fraud. Our focus is on determining the efficacy of each algorithm in different scenarios of fraud detection, understanding their strengths and limitations, and developing a comprehensive approach that optimally combines these methods for robust and reliable fraud detection in the ever-evolving landscape of digital financial transactions. This question is crucial in guiding our research and development efforts towards creating a more secure financial environment in the age of digital banking and e-commerce.

Scope:

Given the short timeframe for our project we intend to focus primarily on the implementation and comparative analysis of the DBSCAN, KMeans, and KNN algorithms in detecting anomalies in a dataset of credit card transactions. Our scope will include data preprocessing, creation of each algorithm from scratch, and evaluation of their effectiveness in identifying fraudulent transactions. However, due to time constraints, we will not delve into extensive fine-tuning of these algorithms or explore additional machine learning techniques beyond the scope of clustering. Instead, we'll prioritise establishing a solid foundation in understanding and applying these specific algorithms, ensuring that our focus remains on delivering tangible and insightful results.

Description:

Our project intends to solve these concerns with the rampant issue of credit card fraud, necessitating a sophisticated solution which needs to be quick and more holistic to account for every aspect of a transaction. The approach begins with the application of KNN, leveraging its ability to discern transaction patterns by analysing the proximity of instances in a feature space. This serves as a foundational element for effective anomaly detection within credit card transactions. Complementing this, K-Means clustering is employed to group transactions based on similarities, thereby enhancing the system's capability to identify and isolate fraudulent activities.

A distinctive feature of this project involves the optimization of the Expectation-Maximization (EM) step within the K-Means algorithm. This step, integral to the refinement of cluster assignments and centroid updates, plays a pivotal role in shaping the overall performance of the algorithm. In the DBSCAN we will also incorporate a BFS and DFS approach to see which one performs better.

Our project adopts a comprehensive strategy that goes beyond mere detection, aiming to strengthen the digital financial landscape. Through the careful application of clustering algorithms and focused optimization efforts, our objective is to achieve prompt recognition and mitigation of credit card fraud. The overarching ambition of our endeavour is to aid in creating robust and secure digital financial frameworks, capable of withstanding the continually changing landscape of cyber threats. This approach is not just about identifying fraud but proactively enhancing the security systems that safeguard digital financial transactions.