## Project Title: Hidden Markov Model for Credit Fraud Detection

Project’s Github: [https://github.com/kunsergio117/CreditFraudDetectionHMM.git]

Team Profile

Team Members:

1. Sergio Gabriel Jiawei Kun

2. Joseph Lee Judkins - [Qualifications and Strengths]

Team Leader: Sergio Gabriel Jiawei Kun

### Project Description

#### 1. Introduction

Credit fraud is a significant and growing issue that impacts individuals and organizations globally. Current systems often struggle to keep up with sophisticated fraudulent activities due to their rigidity and potential inability to detect subtle, evolving patterns of fraud. This project proposes to develop a system using Hidden Markov Models (HMM) to effectively detect credit fraud in transactions. Hidden Markov Models offer a probabilistic framework, possibly capturing temporal patterns in sequential data. This makes it well suited for detecting anomalies in credit card transaction sequences. HMMs can identify deviations by modeling a customer’s typical transaction behavior, which can indicate fraud and other similar fraudulent activities.

#### 2. Problem Diagnosis

The problem domain involves detecting fraudulent transactions in credit systems. The issues with current practices include:

- High False Positive Rates: Existing systems frequently flag legitimate transactions as suspicious because they rely on generalized rules. This is an issue because some customers have erratic behavior, but they are legitimate purchases. False positives result in frustration for customers, especially those whose accounts are frozen unnecessarily, and create burdens for companies or other institutions that must handle increased volumes of fraud investigations.

- Inability to Adapt: Fraudulent schemes are often evolving, with fraudsters constantly innovating ways to scam their victims. Rule based systems struggle to keep pace with them, struggling to adapt and often requiring manual updates. These updates then become more reactionary instead of proactive, which does not help the person who was scammed. These systems lack flexibility to adapt dynamically. This makes them more ineffective when it comes to detecting new behaviors or tactics as patterns evolve further.

- Data Volume: Monetary transactions are the backbone of society, with millions of people making multiple transactions each day. Keeping track of all of them to detect fraud is no small feat. With online banking and e-commerce, these transactions are more virtual than ever, making detection of real-time fraud a massive challenge. Conventional systems can struggle to process this much information at once, and struggle to detect fraud quickly enough to catch fraudsters in the act. This results in delayed responses. With all data needing to be analyzed in real time, the traditional methods can sometimes struggle to keep up with the sheer demand and number of transactions.

- Example scenarios of these issues might include interviews with financial institutions that discuss their current challenges with fraud detection, with those issues being false positives. A global bank might report that while its current fraud detection system effectively blocks fraudulent transactions, the system may also frequently flags legitimate purchases, with this problem being especially prevalent during days like Black Friday, something similar may occur in the case of international travel. The detection system may determine that a purchase was made in another country, and flag it without determining whether the user is in that country on vacation or not. Financial institutions also may express concerns about adaptability, with manual updates being more reactionary than proactive, as it can be hard to guess what a fraudster will do before they do it. In these scenarios, reactionary combat of fraud may not help the people who are first affected by these fraudulent actions. \*

#### 3. Proposed Treatment

To address the diagnosed problems, we propose the following interventions:

- Implementation of HMM that models the sequence of transactions to identify unusual patterns indicative of fraud.

- Transaction data will be ingested by program in the form of csv files (Real-time monitoring functionality can be added in the form of API calls or access to real-time databases).

- User-Friendly Dashboard for analysts to view flagged transactions and insights.

- [https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud] : This Kaggle dataset will be used as our training and testing grounds for our HMM model.

**Metrics for success** will include:

Reduced false positives,

increased detection rates,

and improved analyst response times.

Example scenarios could illustrate how the HMM will flag a suspicious transaction for further investigation.

### Plan for the following weeks

#### 1. Initial Steps:

- Research existing HMM algorithms and their application to fraud detection. Python already has HMM libraries for this purpose. Knowledge of training and testing such a model is necessary.

- [https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud] Testing of the model on this dataset will allow us to score its ability, additionally since it is stated that 492 frauds out of 284,807 transactions are contained allowing us to score it accurately.

- The dashboard, which will provide smoother user-interaction with the application, needs to be done, some front-end frameworks need to be explored such as Flask or Django, or even just a simple table file to be output.

#### 2. Functional Features:

- Transaction Analysis: Detect patterns in transaction sequences by the statistically trained model

- Reporting Tools: Generate reports showing detection rates, false positives, and other relevant metrics. This can be a another text file for debugging and optimization purposes.

Each team member will be responsible for specific functionalities, such as:

- Sergio Kun: Development of the transaction analysis module (for now this is assumed to be done in python as I am most familiar with statistical analysis with it.

- Lee Judkins: Implementation of the alerts generation mechanism (the front-end or user facing segments.

#### 3. Product Ownership:

Each team member’s contributions will be clearly defined to ensure accountability this can be clearly tracked by our github collaboration on the repository.

### Conclusion

This proposal aims to leverage the HMM to provide a robust solution to credit fraud detection. By thoroughly diagnosing the problem, prescribing a targeted treatment, and outlining a clear plan of work, we aim to create a beneficial system for financial institutions.