**A FRAUD DETECTION SOFTWARE USING DEEP LEARNING**

**1.1 Background of the project**

Fraud is an illicit act mainly for financial or personal gain that is constantly evolving with irregularities in its patterns. As more firms venture into digitization, there is more increase in the growth of users in online banking systems which has led to more stolen credit cards credentials and online transaction theft. As a result of this problem, great concern and attention are now drawn to how to prevent and stop the stealing of these credit cards credentials and online transaction theft. Fraud detection is a process or a method used to identify and stop fraudulent activities. A good fraud detection software should spots and stop fraudulent activities at its granular events in real-time before it gets to the transaction phase.

Deep learning is a branch of machine learning that involves training a model with huge amounts of data, it is usually made up of many layers of data processing and can detect patterns in data without being supervised. Technology is constantly changing so as cyber-attacks, therefore governments, organizations, and business owners should be at the forefront and prepare to deal with such unexpected circumstances.

**1.2 Statement of the Problem**

Online transaction fraud is one of the leading cybersecurity problems of our time. This act of deception has resulted in issues such as credit card theft and stolen identity which causes users of online banking systems to lose money and reputational damage. This problem has led to the research and development of so many fraud detection software to help prevent and stop fraudulent transactions. However, those software is inefficient because they involve the traditional manual of feature engineering and detect fraud at the transaction level which cannot detect new patterns of fraudulent attacks. Hence this proposed software will address that issue by identifying and detecting fraud in real-time at its granular events level before reaching the transaction level.

**1.3 Aim and Objectives of the project**

This work aims to build fraud detection software with the working principle of unsupervised machine learning. The following objectives will make sure that the aim is achieved.

1. To pre-process and model transaction data in real-time at its granular events level using the k means clustering algorithm.
2. To further process the data gotten from the level of the granular event with an autoencoder algorithm program using Keras API.

**1.4 Significance of the project**

1. To model at the event level instead of the transaction level which gives us a more granular perspective.

2. To look at both homogenous and heterogenous events that lead to fraudulent transactions.

3. To address the issue of false positive alert transactions

**2.0 Proposed Method**

**2.1 Algorithm techniques and Framework to be used**

1. K-means clustering algorithm

2. An autoencoder deep learning algorithm

3. Keras API, implemented by python.

**2.2 Methods**

The first part of the fraud detection software will consist of the k means clustering algorithm which will be responsible for pre-processing the user's profile that is initiating the transaction. The k- means clustering algorithm will try to group the data into clusters using identified similarities, afterward, the grouped clusters will be further passed to the second part of the software which will be made up of the autoencoder deep learning neural network for further processing. These data will be converted from categorical values into numerical forms by a method called feature encoding through the first layer of the autoencoder neural network, this is because statistical models work best with numerical values than categorical values. This layer is responsible for receiving the grouped cluster data as inputs after then the encoded data will be passed to the hidden layers for further processing to get the actual and the most information needed based on fraudulent and nonfraudulent features. The hidden layers for this software are made up of two layers, these are to ensure proper cleansing and processing of the data while in an encoded format after which the final layer receives the processed data and reconstructs the data according to their input representations. These output results will determine the kind of data being channeled through the input layers whether fraudulent or nonfraudulent transactions because the neural network will only produce a pattern of data based on normal transactions that it has been trained for, otherwise, it will reproduce a different type of data pattern if the transaction data are not in these case a normal transaction type but a fraudulent type. The role of the k- means clustering algorithm is to group transaction user credentials at its granular events stage and cluster them based on similar attributes the autoencoder takes over by learning deeply and identifying the most necessary and important data point that truly stands for either fraudulent or nonfraudulent transactions.

Keras model will be used to implement the autoencoder algorithm. The design will have five layers, one input layer, two hidden encoded layers, and two decoded output layers.

ESSENTIAL DATA FEATURE SELECTION AND CLUSTERING BY K-MEANS ALGORITM.

**SOFTWARE ARCHITECTURE AND WORKFLOW PARTS**

DATA ACQUISATION PHASE AT GRANULAR EVENTS.

FURTHER DATA PROCESSING AND OUTPUT RESULT BY AN AUTO ENCODER NEURAL NETWORK.