Name: Meti Kebede

Project Name: Credit card Fraud Detection

# Introduction

The credit card Fraud detection dataset is available on Kaggle and is widely used for data mining and machine learning projects. This dataset contains transactional data of credit card usage by European cardholders in September 2013, spanning a duration of two days. During this period, 492 fraudulent transactions were detected, out of a total of 284,807 transactions. The purpose of this white paper is to offer an overview of data mining techniques applied to the credit card fraud detection dataset, encompassing various aspects such as data preparation, exploratory data analysis, modeling, and evaluation.

# Dataset:

The credit card fraud dataset is available in CSV format and can be downloaded from Kaggle or OpenML.org.

# 3.Data Preparation:

Once the data is downloaded, the initial step is to prepare it, which involves identifying and addressing missing values and outliers. This stage includes data cleaning techniques such as removal or addition of values as appropriate. Subsequently, the data is transformed into a format that is suitable for analysis.

# Exploratory Data Analysis:

The next step in will be performing Exploratory Data Analysis (EDA). This step involves the process of visualizing and summarizing the data to obtain valuable insights into the interrelationships between the variables. For this specific dataset, a correlation matrix, histograms, and scatter plots will be utilized to establish and evaluate the connections among different variables in the dataset.

# Modeling

This stage involves the selection of appropriate machine learning algorithms. In this project Random Forest, Decision tree and Support Vector machine will be used. The goal of modeling here is to detect fraudulent transaction and classify them accordingly. In order to build the model, the dataset will be divided into two main groups. One for the training the model and the other for testing the trained model’s performance.

# Evaluation

The evaluation of the models is a crucial component of this project. This phase involves the calculation of various metrics such as accuracy, precision, recall, and F1 score, which determine the effectiveness of the models in detecting fraudulent transactions. The model that exhibits the highest accuracy and is deemed most appropriate for handling this specific dataset will be selected for further use.