CREDIT CARD FRAUD DETECTION

**Aims and Objectives:**

**This is aims to develop a Machine Learning model that for credit card fraud detection from given.**

**Objectives:**

**• Converting data into an appropriate form using various preprocessing techniques for the the implementation of Machine Learning algorithms.**

**Machine Learning:**

**Machine Learning is the area of study which enables machines to learn without being explicitly programmed. Machine Learning is defined as the computer program learns from experience E with respect to some class of tasks T and performance measure P when its performance at tasks in T, as measured by P, strengthens.**

**Supervised Learning:**

**The most popular model for performing Machine Learning processes is supervised learning. It is commonly used for data where the mapping between input-output data is accurate. Supervisedlearning is the subset of Machine Learning which concentrates on learning a model of classification or regression, that is, learning from labeled test data.**

**Unsupervised Learning:**

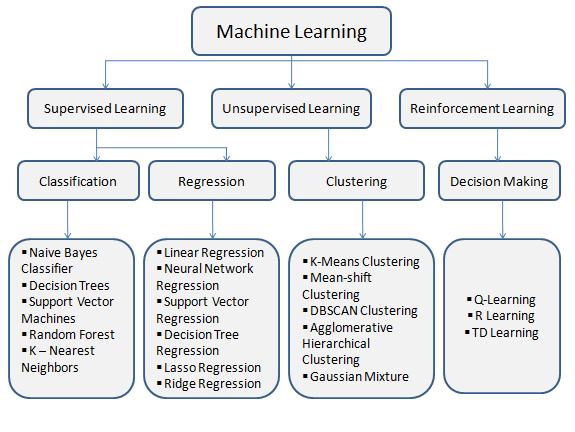
**The data is not explicitly labeled into different classes in the case of unsupervised learningthat is there is only unlabeled data. By identifying implicit patterns the model can learn from the data.**

**Unsupervised Learning categorizes the densities, structures, related segments, and other similar properties based on the data.**

**Reinforcement Learning:**

**Reinforcement Learning is a sub-field of Machine Learning. In a given scenario, it is about taking the best possible action or path it will follow in a specific scenario.**

**Data Collection: Gather historical data on product sales, including variables like time, price, marketing efforts, and external factors (e.g., holidays, economic indicators).**

****

**Program:**

**import pandas as pd**

**from sklearn.model\_selection import train\_test\_split**

**from sklearn.ensemble import IsolationForest**

**from sklearn.metrics import classification\_report,**

**accuracy\_score**

**data = pd.read\_csv(‘credit\_card\_data.csv’)**

**X = data.drop(‘Class’,axis=1)**

**y = data[‘Class’]**

**X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y,test\_size=0.2, random\_state=42)**

**model=IsolationForest(contamination=0.01,random\_state=42)**

**model.fit(X\_train)**

**y\_pred = model.predict(X\_test)**

**y\_pred[y\_pred == 1] = 0**

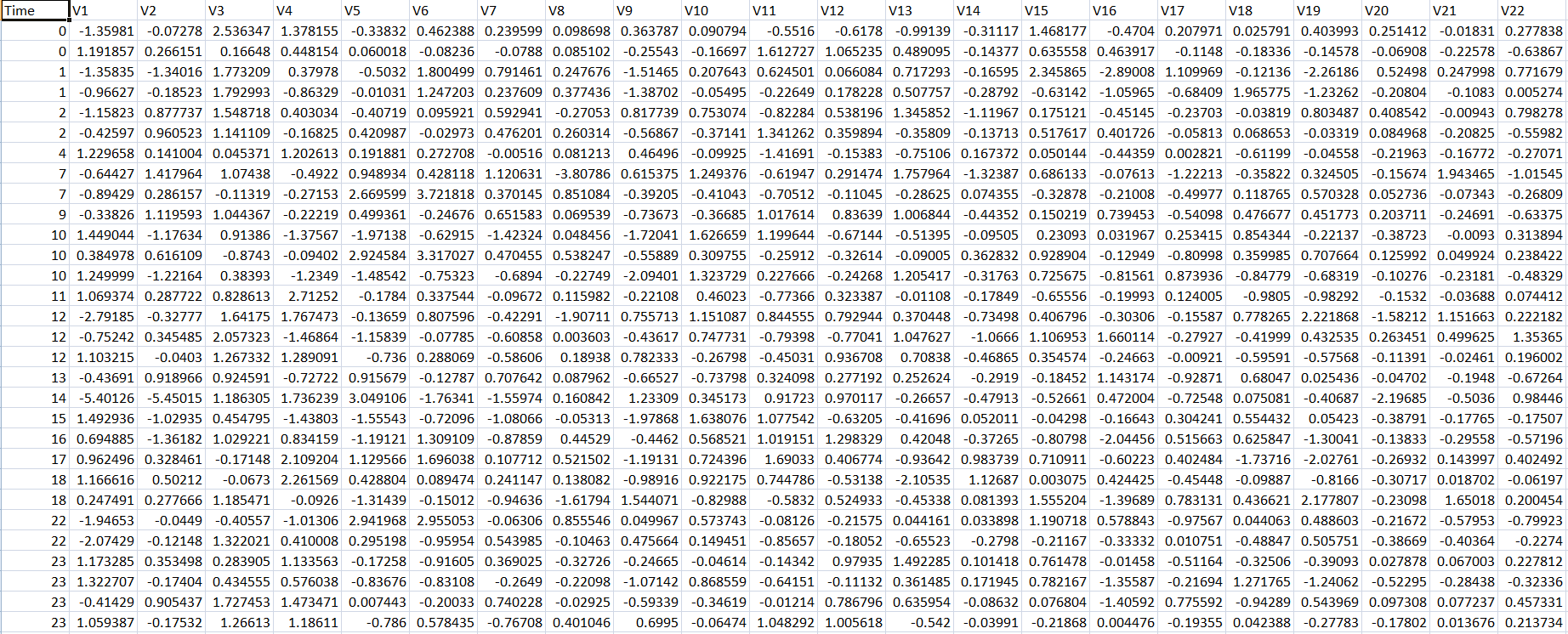
**y\_pred[y\_pred == -1] = 1**

**accuracy = accuracy\_score(y\_test, y\_pred)**

**print(&quot;Accuracy: {:.2f}%&quot;.format(accuracy \* 100))**

**print(classification\_report(y\_test, y\_pred))**

**(**[**https://skilluptech-my.sharepoint.com/:b:/g/personal/simrang\_skillup\_online/EZjznEGC5VVAvWNefQ-uERkBeFLCyboV4mAuleYE3ENxLA?e=2fqSfm**](https://skilluptech-my.sharepoint.com/:b:/g/personal/simrang_skillup_online/EZjznEGC5VVAvWNefQ-uERkBeFLCyboV4mAuleYE3ENxLA?e=2fqSfm)**)**

****

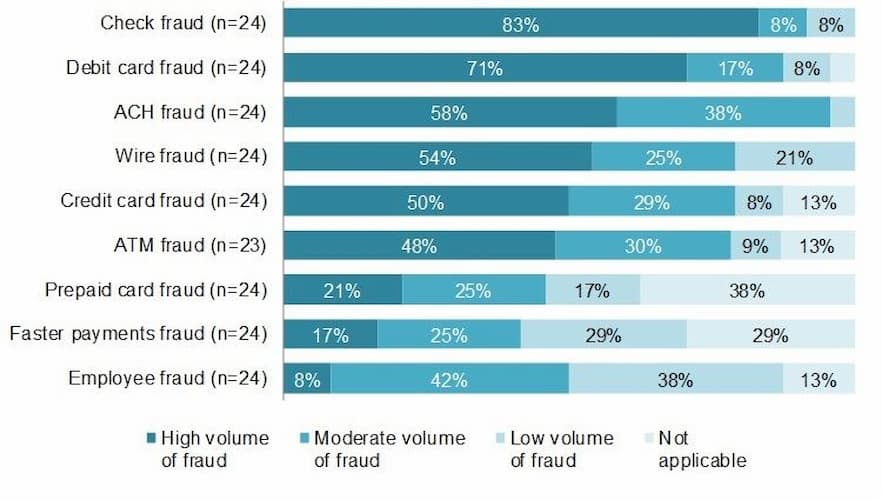
**The Given dataset contain huge amount of data so we can’t present the full**

**Dataset on the Document.**

**So we present the sample of graph of the about dataset.**

**Feature Importance:**

**The Credit Card Fraud Detection Problem includes modeling past credit card transactions with the knowledge of the ones that turned out to be fraud. This model is then used to identify whether a new transaction is fraudulent or not.**

****

**Data preprocessing:**

**Before applying Machine Learning algorithms some of the missing values have been found which can impact the model’s output so this should be handled. The ’item weight’ and ’outlet size attributes have 17 percent, and there is 28 percent of missing values.**

**Data Cleaning:**

**This involves identifying and correcting errors or inconsistencies in the data, such as missing values, outliers, and duplicates. Various techniques can be used for data cleaning, such as imputation, removal, and transformation.**

**Data Integration:**

**This involves combining data from multiple sources to create a unified dataset. Data integration can be challenging as it requires handling data with different formats, structures, and semantics. Techniques such as record linkage and data fusion can be used for data integration.**

**Data Transformation:**

**This involves converting the data into a suitable format for analysis. Common techniques used in data transformation include normalization, standardization, and discretization. Normalization is used to scale the data to a common range, while standardization is used to transform the data to have zero mean and unit variance. Discretization is used to convert continuous data into discrete categories.**

**Data Reduction:**

**This involves reducing the size of the dataset while preserving the important information. Data reduction can be achieved through techniques such as feature selection and feature extraction. Feature selection involves selecting a subset of relevant features from the dataset, while feature extraction involves transforming the data into a lower-dimensional space while preserving the important information.**

**OUTPUT:**

