# **CREDITCARD DATASET**

Official transaction dataset on European cardholders across the month of September 2013, based-on online-offline channels. It is good and meant for machine learning projects in audience detection against fraud transaction activities.

The crediteard.csv dataset contains details of 284,807 transactions out of which 492 are fraudulent transactions.

The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.



# **FEATURES:**

### 1. Time:

Type: Numeric

**Description:** The number of seconds since the first transaction was recorded in the dataset.

### 2. V1 to V28:

Type: Numeric

**Description:** These are the main components derived from PCA (Principal Component Analysis), which was done to mask original features. There are such 28 features in total in the dataset, indexed V1 to V28.

# 3. Amount:

Type: Numeric

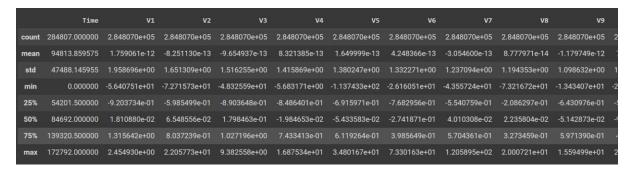
**Description:** Values represent the amount spent during the transaction and differ greatly with some being very small and others large.

#### 4. Class:

**Type:** Categorical (0-1)

**Description:** The target variable of Class 0 and Class 1. The Class 1 value indicates that a transaction is a fraud, while a Class 0 value indicates that it is a non-fraudulent transaction.

### **SUMMARY STATISTICS:**



**Total Transaction Count:** The number of different transactions found in the dataset.

**Fraud Transactions:** The total number of transactions that have been classified as fraudulent (1).

**Non-Fraudulent Transactions:** The total number of transactions classified as normal (0).

# **CLASS IMBALANCE:**

In general, there is an imbalance. Only a very small portion of such transactions can be considered fraudulent when placed alongside normal ones. That would present a set of problems when it comes to model training and evaluation.

### **USE CASES:**

**Fraud detection:** Designing a model that identifies fraudulent transactions based on generic transaction characteristics.

**Anomaly Detections:** Analysis of the transaction data patterns to detect anomalies.

### **CONCLUSION:**

The creditcard.csv dataset provides a fine staging ground for the project launch and training of models and the prototype testing of their performance in fraud detection. It is important to take into account the features and their meanings in order to provide strong footing for model development and analysis.