**Credit Card Fraud Detection**

**Anonymized credit card transactions labeled as fraudulent or genuine**

From: <https://www.kaggle.com/mlg-ulb/creditcardfraud>

**Main objective:**

Recognize fraudulent credit card transactions. CLASSIFICATION PROBLEM.

**Dataset:**

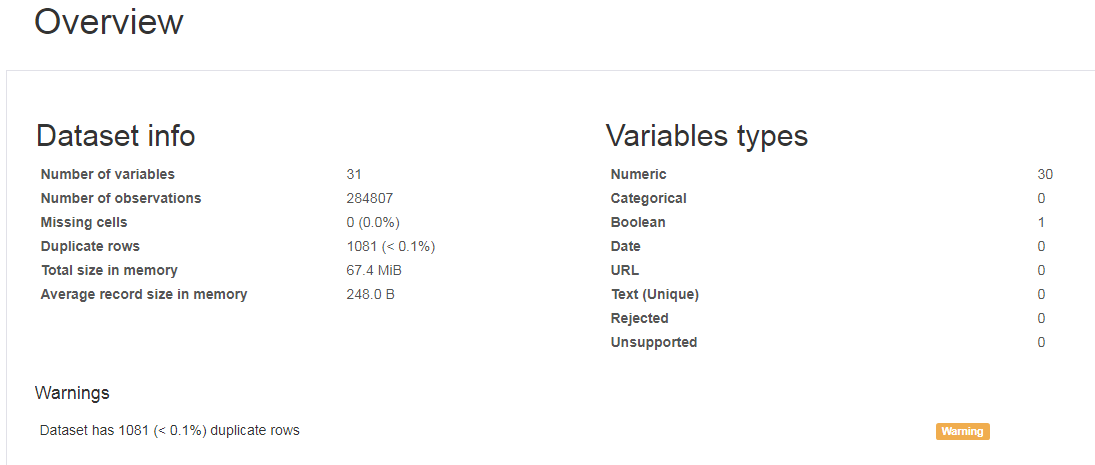
The datasets contains transactions made by credit cards in September 2013 by european cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

Column names:

'Time', 'V1', 'V2', 'V3', 'V4', 'V5', 'V6', 'V7', 'V8', 'V9', 'V10', 'V11', 'V12', 'V13', 'V14', 'V15', 'V16', 'V17', 'V18', 'V19', 'V20', 'V21', 'V22', 'V23', 'V24', 'V25', 'V26', 'V27', 'V28', 'Amount', 'Class'

* **V1-V28**: due to confidentiality issues, they could not provide the original features and more background information about the data. Features V1, V2, ... V28 are the principal components obtained with PCA.
* **Time:** contains the seconds elapsed between each transaction and the first transaction in the dataset.
* **Amount:** is the transaction amount, this feature can be used for example-dependant cost-senstive learning.
* **Class:** is the response variable and it takes value 1 in case of fraud and 0 otherwise.

Using panda’s profile library we can extract some information easily.



Correlation matrix:

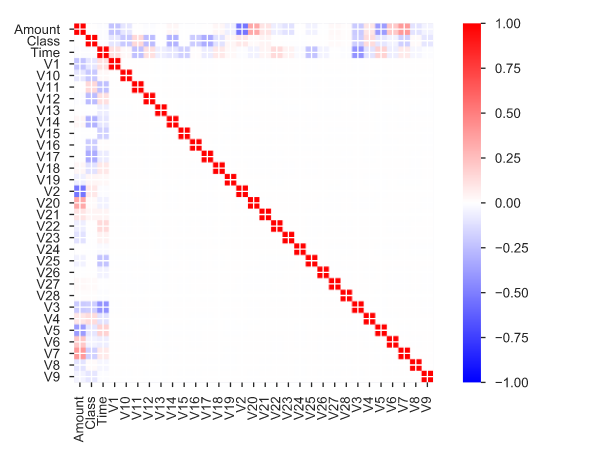


Figure 1. Correlation Matrix

We can see the main correlated variables:

V2 - Amount (negative correlation)

V5 - Amount (negative correlation)

V7 - Amount