# Data Understanding

The data for this project comes from [Kaggle](#) and was posted by the machine learning group at the university of brussels. This is the only data source for this project. It comes in a .csv format and the size of it is 144 megabytes. To obtain the data for yourself simply click the link at the bottom of this page, navigate to the upper right side of the webpage that you are directed to, and click the white/black download button.
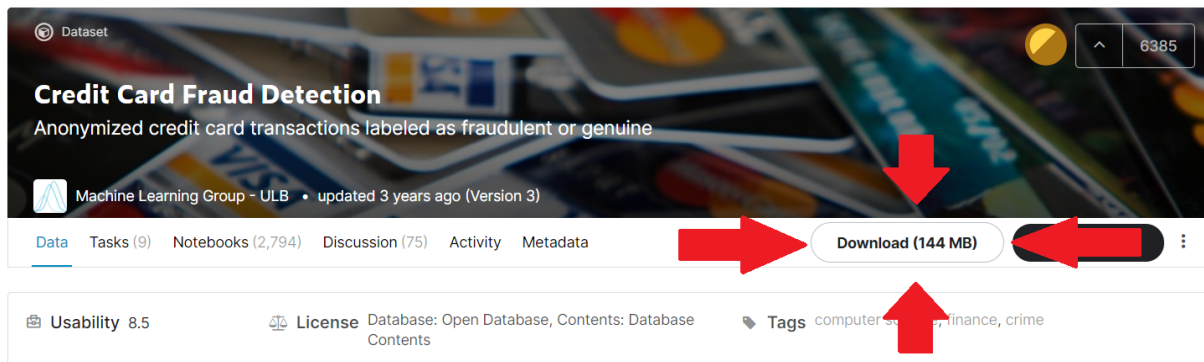
"The dataset contains transactions made by credit cards in September 2013 by european cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions."[1]

The data has 30 features and 1 target variable. The target variable is a 1 if the transaction is known to be fraud, and a zero if it is legitimate. 28 of the features have been anonymized through principal component analysis to protect cardholder information.[1] All that is known is that these features represent information about the transaction. We can infer that the location, vendor information, and account information such as credit score are included. Along with these 28 anonymous features we are given the amount spent in the transaction (presumably in euros) and the time of the transaction, which is measured in seconds since the first transaction.[1]

Sources

1
Credit Card Fraud Detection
Machine Learning Group ULB
https://www.kaggle.com/mlg-ulb/creditcardfraud

Acknowledgements