**1 - Brain Cancer Gene Expression Classifier - CuMiDa**

[Curated Microarray Database (CuMiDa)](http://sbcb.inf.ufrgs.br/cumida), a repository containing 78 handpicked cancer microarray datasets, extensively curated from 30.000 studies from the Gene Expression Omnibus (GEO), solely for machine learning. Use a Deep Neural Network to classify cancers from genes and use Principal Component Analysis to identify key genes.

Bruno Grisci - <https://www.kaggle.com/brunogrisci/brain-cancer-gene-expression-cumida>

**2 - Religious and Spiritual Texts NLP**

Word cloud analysis, Topic modeling, Word & document frequencies, Charts of the highest tf-idf words in each text within the corpus, Word dendrogram cluster, Pairwise correlation, Sentiment analysis, Word Network.

Dan Pfeiffer - <https://www.kaggle.com/metron/public-files-of-religious-and-spiritual-texts>

**3 - Credit Card Fraud Detection**

Identify fraudulent credit card transactions from a highly unbalanced dataset with .172% positive label class.

Machine Learning Group - <https://www.kaggle.com/mlg-ulb/creditcardfraud>