**The Depression Dataset**

Depression is a severe mental disorder with characteristic symptoms like sadness, the feeling of emptiness, anxiety and sleep disturbance, as well as general loss of initiative and interest in activities. Additionally, features like the feeling of guilt or worthlessness, reduced energy, concentration problems, suicidality and psychotic symptoms might be present. The severity of a depression is determined by the quantity of symptoms, their seriousness and duration, as well as the consequences on social and occupational function. Depressions are also common in Bipolar disorder, another severe psychiatric disorder. The main difference between unipolar depression and bipolar disorder is the periodic occurrence of mania in the latter, a state associated with inﬂated self-esteem, impulsivity, increased activity, reduced sleep and goal-directed actions. Both diseases are genetic disorders, and might be understood as a genetic vulnerability to the environment disturbing the internal biological state and potentially trigger mood episodes. Depression is associated with disrupted biological rhythms caused by environmental disturbance like seasonal change in daylight, alteration of social rhythms due to for instance shift-work or longitude traveling; besides linked to lifestyles associated with diurnal rhythms inconsistent with the natural daylight cycle. The appearance of depressive symptoms related furthermore to physical health issues, medical side effects, life events and social factors, besides alcohol and substance abuse, and such factors might also potentially cause symptoms of depression in all humans. The global lifetime prevalence of depression is roughly 15%, but the incidences of episodes with a severity level not meeting the requirements for a depressive diagnosis are far more prevalent. Actigraph recordings of motor activity are considered an objective method for observing depression, although this topic is far from exhaustively studied within psychiatric research.

### **Content**

The dataset contains two folders, whereas one contains the data for the controls and one for the condition group. For each patient, a csv file has been provided containing the actigraph data collected over time. The columns are: timestamp (one minute intervals), date (date of measurement), activity (activity measurement from the actigraph watch). In addition, the MADRS scores are provided in the file "scores.csv". It contains the following columns; number (patient identifier), days (number of days of measurements), gender (1 or 2 for female or male), age (age in age groups), afftype (1: bipolar II, 2: unipolar depressive, 3: bipolar I), melanch (1: melancholia, 2: no melancholia), inpatient (1: inpatient, 2: outpatient), edu (education grouped in years), marriage (1: married or cohabiting, 2: single), work (1: working or studying, 2: unemployed/sick leave/pension), madrs1 (MADRS score when measurement started), madrs2 (MADRS when measurement stopped).

More : <https://datasets.simula.no/depresjon/#dataset-details>

### **Starter Code**

* [@docxian](https://www.kaggle.com/docxian) :  
  [Evaluate motor activity of depression patients (and healthy control group)](https://www.kaggle.com/datasets/arashnic/the-depression-dataset?resource=download)

### **Acknowledgements**

Enrique Garcia-Ceja, Michael Riegler, Petter Jakobsen, Jim Tørresen, Tine Nordgreen, Ketil J. Oedegaard, Ole Bernt Fasmer, Depresjon: A Motor Activity Database of Depression Episodes in Unipolar and Bipolar Patients, In MMSys'18 Proceedings of the 9th ACM on Multimedia Systems Conference, Amsterdam, The Netherlands, June 12 - 15, 2018.

### **Inspiration**

* The available data may eventually help researchers to develop systems capable of automatically detecting depression states based on sensor data. This dataset can be suitable (but not limited to) for the following applications:
  + (i) Use machine learning for depression states classification
  + (ii) MADRS score prediction based on motor activity data
  + (iii) Sleep pattern analysis of depression v.s. non-depressed participants

This dataset can be used as the basis for evaluating different machine learning methods and approaches such as: cost-sensitive classification and oversampling techniques for imbalanced class problems. This dataset is also suitable for comparing different machine learning classification approaches such as feature based and deep learning based methods like convolutional neural networks and recurrent neural networks for time series.