Schizophrenia and Bipolar Disorder Classification

## Problem Statement:

using ML tools to classify between Schizophrenia and Bipolar Disorder using the dataset attached below.

The dataset is in the form of a pickle file (dictionary file).

It has 6 feature keys - ALFF, fALFF, .... - and for each feature, 14 different atlas keys - AAL, Power, .....

## Task:

For your task, you'll be using only the ReHo feature key (Regional Homogenity), and its 14 corresponding atlas keys to classify between Schizophrenia and Bipolar Disorder with >70% accuracy.



Figure : Distribution of mental disorder

# Approach

|  |  |  |  |
| --- | --- | --- | --- |
| Number | Task | Concepts/ Ideas used/ library used | Conclusion |
| 1 | read data | Read data using a custom function, pandas, and numpy | Acceleration and Rotation Magnitude data for walking and running is loaded. |
| 2 | exploratory data analysis & visualize data | Matplotlib, seaborn | Plot weight height age and Gender distribution. |
| 3 | convert to time series | datetime | Datetime variable created |
| 4 | perform ordinary least squares | Fitting a model in statsmodels typically involves 3 easy steps:  Use the model class to describe the model  Fit the model using a class method  Inspect the results using a summary method  statsmodel | It is seen that r-squared is 0.509, 50.9% of the variance in the outcome variable is explained by the model  Model variance is not explained by these variables. |
| 5 | perform time series decomposition of model | statsmodel.tsa.seasonal.seasonal\_decompose, pandas, matplotlib | Removed the seasonality and trend. |
| 6 | observe seasonality between walk and run. (act = 1 or 0) | Matplotlib, Tableau dashboard. | Looking at the repeatability between 100 to 350 observations for run and walk |
| Here we understand that there is a seasonlity between 100 observations. This means 2 seconds for 50Hz observations. | | | |
| 7 | again perform ols | numpy to reshape data, and target variables | r-squared value improved from 0.509 to 0.601.  new features can explain variability in the dataset much better. |
| 8 | Create fully connected neural (FCN) network function | Keras, Sequential, Dense, Dropout, SGD | Perform cross validation on Sequential model, for binary classification. |
| 9 | Perform Strafied K Fold CV | sklearn, pipeline, Kerasclassifier, standardscalar, strafiedKFold cross-validation | Mean Accuracy is 94.77% and standard deviation is 0.57% |
| 10 | Perform Classification on test set | Confusion\_matrix, print\_confusion\_matrix | Print accuracy score, and confusion matrix |
| 11 | Print model accuracy metrics | r2\_score,classification\_report, accuracy\_score, roc\_curve, roc\_auc\_score, auc | Print model classification report with F1, recall and precision. |

# Conclusion:

A classifier for a hand activity recognition is existing

The classifier can predict whether the user is walking or running based on the magnitude of userAcceleration data.

## Start:

This model used EDA(exploratory data analysis) and OLS(ordinary Least Squares) method for classifying the output parameter and trying to determine the most important features needed in the model.

This helps us understand that seasonality can be captured when observations of a feature over a period of time are used for classification.

For the current scenario we need to classify between two classes walk or run, hence we have taken only one (userAcceleration Magnitude) variable and a period of time as 100ms.

Create a Sequential FCN with 1 hidden layer and 1 dropout layer model so precision accuracy and recall for walk and run are above 95%.

## Conclusion:

A simple 1 hidden layer FCN model with input single variable is a good model.

## Drawback:

It needs 2 seconds data. Thus we need time before we can make prediction.

The model can be improved further by adding one more feature ().