Phase2 – Feature Selection

in the previous part we tried to extract features from data, after extracting the features, in this step we try to select better features, to do this, we use different components of feature selection methods:

- How well each feature acts on its own
- How different is each feature from the rest of features in general

- First (How well each feature acts on its own)

  We use Decision Tree and try to classify data with only one feature to check how well a feature can differentiate between data with accuracy score

  - Time Domain Features

```
Time Domain Features :

Peek To Peek(PTP) accuracy score : 0.95
AASS accuracy score : 0.9
Singular Spectrum Analysis(SSA) accuracy score : 0.66
Log Detetct(LD) accuracy score : 0.96
Zero Crossings(ZC) accuracy score : 0.68
```

  - Statistical Features

```
Statistical Features :

Mean accuracy score : 0.71
Median accuracy score : 0.97
Percentile accuracy score : 0.95
Standard Derivation(STD) accuracy score : 0.96
Histogram accuracy score : 0.95
```

  - Entropy Features

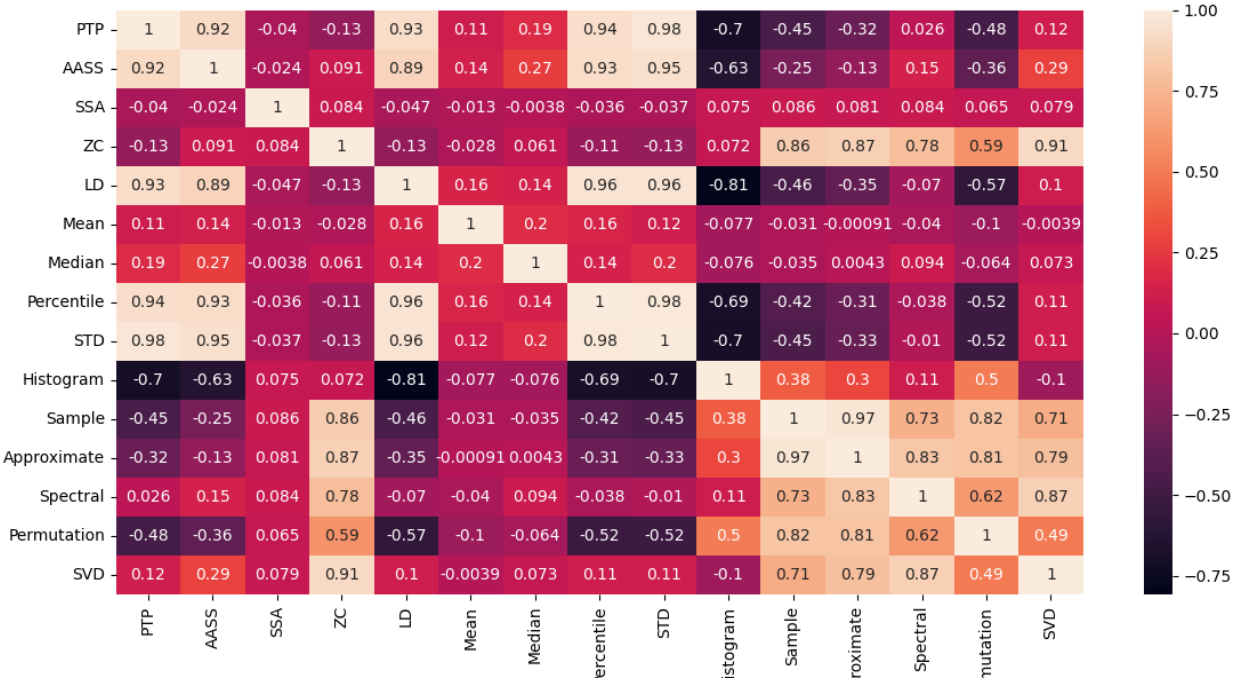
```
Entropy Features :

Sample Entropy accuracy score : 0.79
Approximate Entropy accuracy score : 0.65
Spectral Entropy accuracy score : 0.6
Permutation Entropy accuracy score : 0.8
SVD Entropy accuracy score : 0.68
```

- Second (How different is each feature from the rest of features in general)
  correlation is a well-known similarity measures between two features. If two features are linearly dependent, then their correlation coefficient id ±1. If the features are uncorrelated, the correlation coefficient is 0.
  The correlation matrix is a square matrix that contains the Pearson product-moment correlation coefficient.

Features with high correlation are more linearly dependent and hence have almost the same effect on the dependent variable. So, when two features have high correlation, we can drop one of the two features.

Negative correlation is a relationship between two variables in which one variable increases as the other decreases, and vice versa.

- Correlation between all features



- For Normalization, we divide the absolute value of Correlation by 10.
- For combining two features, we use the same method as f1-score.

$$\text{F1-score} = \frac{\frac{1}{normalizedCorrelation} \times accuracy}{\frac{1}{normalizedCorrelation} + accuracy}$$

Features with higher f1-scores are selected.

After iterating over all features and calculating f1-score, we classify data with selected features using decision tree.

- Top 5 features

```
selected features
Median
LD
STD
PTP
Percentile

Classifying datas with top 5 features :
Cross Validation Score 0.958
Accuracy : 0.95
```

- Top 8 features

```
selected features
Median
LD
STD
PTP
Percentile
Histogram
AASS
Permutation

Classifying datas with top 8 features :
Cross Validation Score 0.9640000000000001
Accuracy : 0.98
```

- Top 10 features

```
selected features
Median
LD
STD
PTP
Percentile
Histogram
AASS
Permutation
Sample
Mean

Classifying datas with top 10 features :
Cross Validation Score 0.958
Accuracy : 0.98
```

- So top 8 features are selected. Due to the type of data, we have, statistical and time domain features are chosen more.

Time domain features

Peek To Peek: shows rising and falling of signal.
Log Detect: this feature particularly used in pulse detection which is highly related to our dataset.
AASS: calculated vertical length of a signal.

Statistical features

Median: median is the data point that separates the upper half of the data values from the lower half.
Standard derivation: shows variation or dispersion of dataset.
Percentile: Percentile divides a data set into the 100 equal parts. A percentile is a measurement that tells us what percent of the total frequency of a data set was at or below that measure. (We use 80 for measure)
Histogram: It shows the frequency of data distribution.

Entropy features

Permutation: it provides a quantification measure of the complexity of a dynamic system by capturing the order relations between values of a time series.

Features that provide new information about the data are selected, and features that have duplicate information are not selected, such as mean or like Signal Spectrum Analysis (SSA) which is similar to AASS.