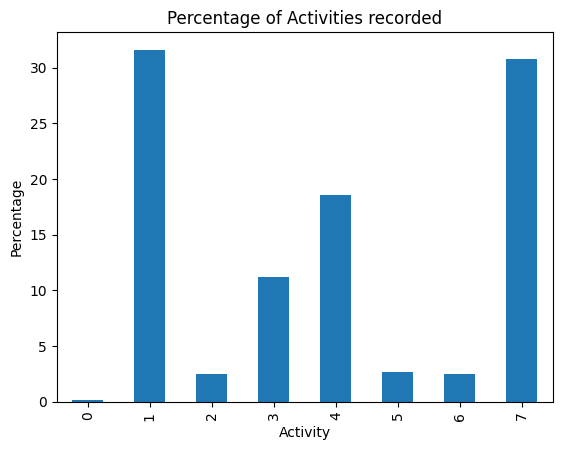
Human Activity Prediction using Accelerometer data

Overview

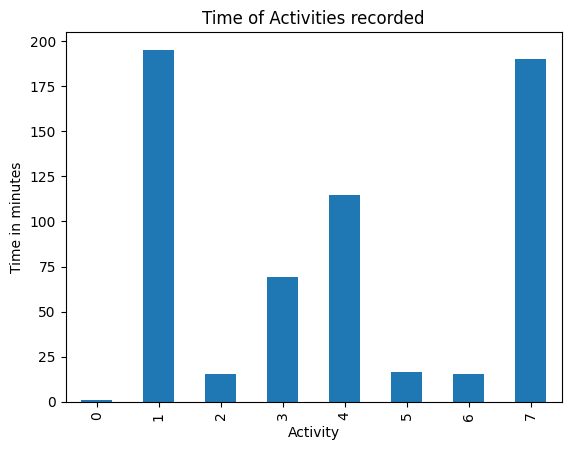
To develop a machine learning model to predict Human activities based on the accelerator meter data of 15 participants with a sampling rate 52 Hz and overlapping of 50%.

Exploratory Data Analysis

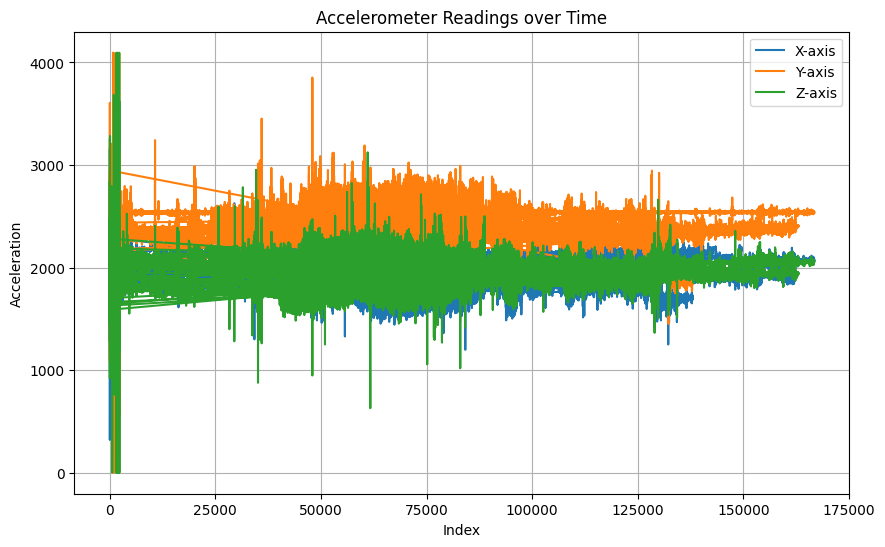
1. Given dataset contains Time Index, Time Domain Signal data (x, y and z acceleration) and acceleration outcome as label
2. No NULL values in the dataset.
3. Below are the label details.
   * 1. 0 - No activity
     2. 1 - Working at Computer
     3. 2 - Standing Up, Walking and Going up\down stairs
     4. 3 - Standing
     5. 4 - Walking
     6. 5 - Going Up\Down Stairs
     7. 6 - Walking and Talking with Someone
     8. 7 - Talking while Standing
4. As per the given data activity 0 – No activity is performed less and activities 1 – Working at Computer and 7 – Talking with Standing performed more by all participants together. As per the amount of data gathered activity 0 can be ignored.



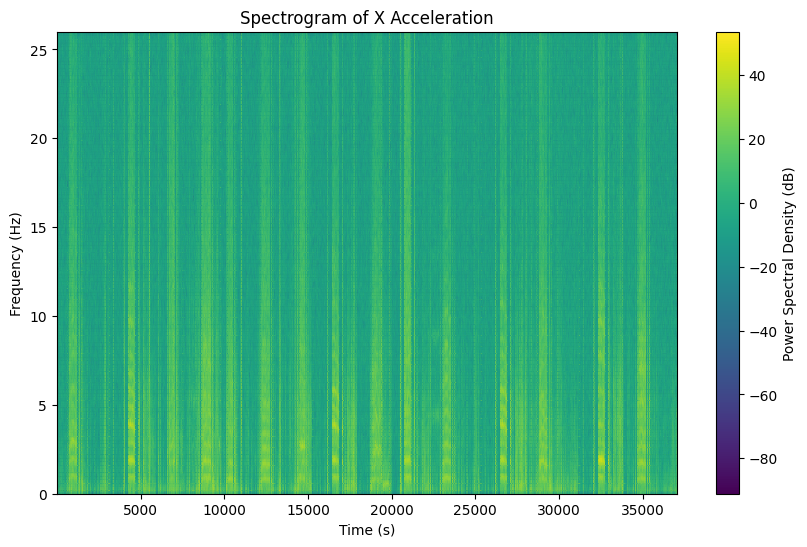
1. Activities 2 – Standing Up, Walking and Going up\down stairs 5 – Going Up\Down Stairs and 6 - Walking and Talking with Someone are performed very less compared to Activities 1 and 7. It clearly creates a class imbalance.



1. All participants started with activity 1 – Working at Computer. By plotting the Accelerometer x, y and z Readings and Time Index, we can see an unusual accelerator recordings at the starting of all participants (i.e. starting readings of activity 1 is seems like outliers and to be handled). We can consider this as outliers.



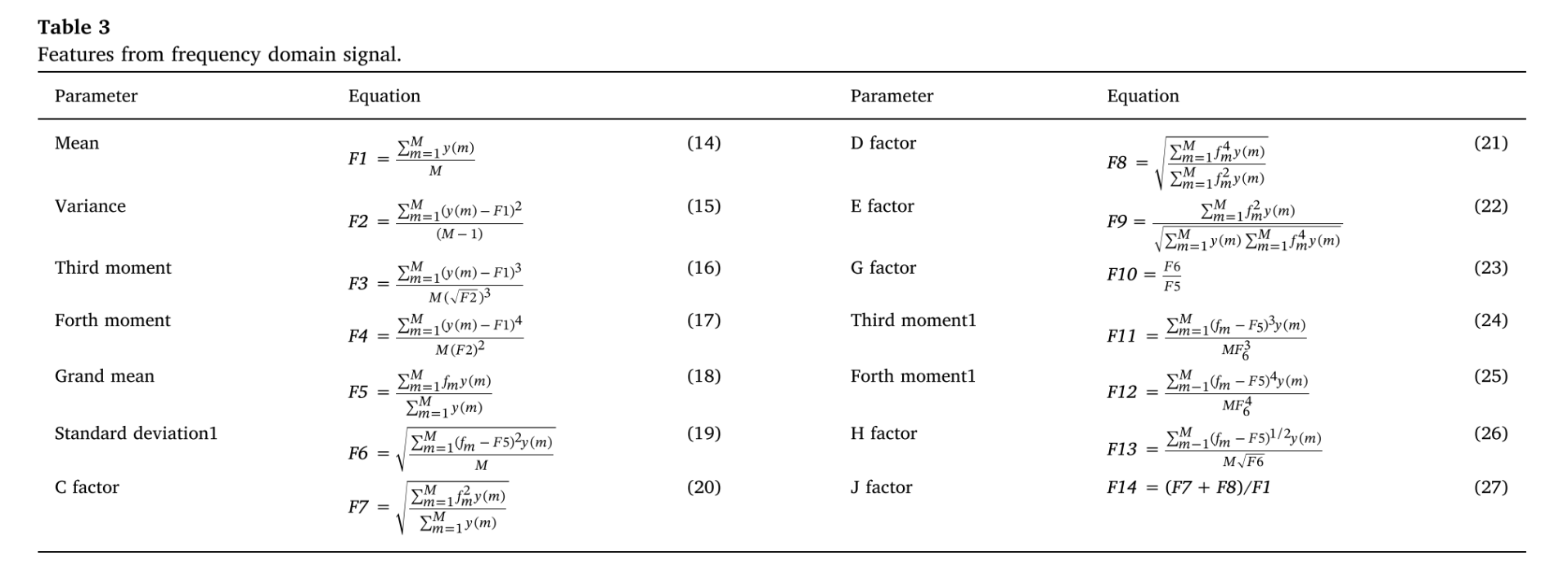
1. We are generating the Frequency Domain Signals using Time Domain Signals dataset by applying Fast Fourier Transform method. We can also analyze the properties of signals using Frequency domain signals for additional information.
2. As per the Power Spectral Density analysis, we can see the energy concentration based on time which provides insights about the peaks in Amplitude vs Frequency Plot. (Peaks highlighted in Yellow which denotes the high energy concentration)

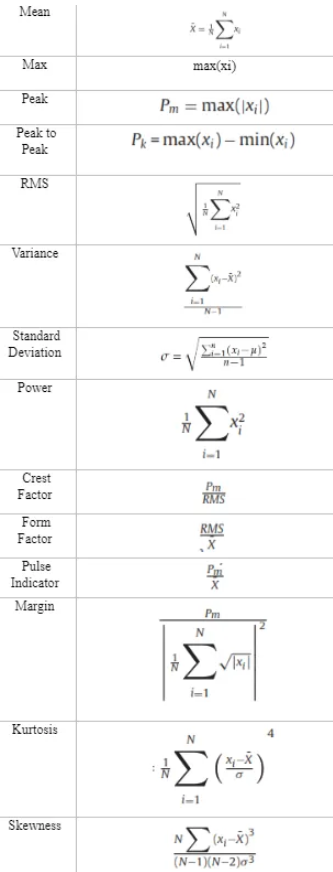
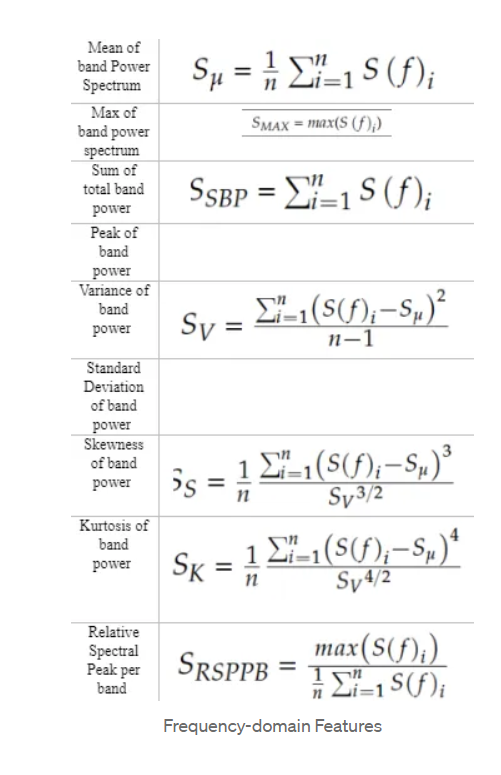


Feature Extraction

<https://link.springer.com/article/10.1007/s12652-018-1110-y>

1. Extracting Features from signals consists of three parts. Statistical, Temporal and Spectral.
2. Extracted Statistical features like Magnitude of Signals, Mean, Median, Skewness, Kurtosis, 10th, 25th, 50th, 75th and 90th percentile, Minimum value, Maximum value, Variance, Standard Deviation.
3. Using Fast Fourier Transform, converted the Time Domain Signal to Frequency Domain Signal and extracted features like Mean, Minimum value, Maximum value, Mean Power, Minimum Power, Maximum Power.



Model Selection

1. As the paper suggests, tried with RandomForestClassifier and results looks good. Achieved an Average accuracy of >88% and F1 score >88% using cross-fold validation with the all the cases included.
2. Created different models as below
   1. After removing No Activity (0) – Accuracy >88% and F1 score > 97%
   2. After removing Outliers – Accuracy >91% and F1 score > 98%
   3. After removing Activities (0, 2 & 6) – Accuracy >93%

Future Works

1. Identifying and Removing NOISE from signals
2. Furthermore analysis on Feature Selection and Generation
3. Trying with other ML & DL algorithms