**MLHack\_01: Fall Detection**

SUbmitted ON: 29th NOV 2019 Anand Kumar Singh

Contents

[Objective 2](#_Toc25920087)

[Scope 2](#_Toc25920088)

[1.1. In Scope 2](#_Toc25920089)

[1.2. Out of Scope 2](#_Toc25920090)

[2. Approach 2](#_Toc25920091)

[2.1 Flow Diagram 3](#_Toc25920092)

[3. Used Classification/Clustering algorithm 3](#_Toc25920093)

[4. Execution steps 4](#_Toc25920094)

[4.1. Step 1: Loading of data into Python 4](#_Toc25920095)

[4.2. Step 2: Baseline model 4](#_Toc25920096)

[4.3. Step 3: Data pre-processing for final model 5](#_Toc25920097)

[4.4. Step 6: Retune the Model 5](#_Toc25920098)

[5. EDA Analysis 5](#_Toc25920099)

[6. Output and Recommendation 5](#_Toc25920100)

# Objective

We trying to build automated fall detection model with help of fitted motion sensor record inside the subjected human body. We have six independent variables (TIME, SL, EEG,BP,HR,CIRCLUATION)

And one dependent variable (ACTIVITY) which have 6 categorical data as below

0🡪Is stand for “**Standing**”

1🡪Is stand for “**Walking**”

2🡪Is stand for “**Sitting**”

3🡪Is stand for “**Falling**”

4🡪Is stand for “**Cramps**”

5🡪Is stand for “**Running**”

Brif about All variables

**Time :** Monitoring Time

**SL**: Sugar Level

**EEG**: Electroencephalograph Monitoring Rate

**BP**: Blood Preasure

**HR**: Heart beat rate

**Circulation**: Blood Circulation

This Sensor device sends all the above mentioned six different positions in real time every 4 seconds with the help of IOT device. We apply statistical analysis and machine learning algorithms to find out insights and predict the fall detection.

# Scope

## In Scope

1. Loading of data, cleaning and processing the data for analysis
2. EDA Analysis
3. Find out best suited Relationship between Fall activity with other given features in dataset.

## Out of Scope

1. Dependent variable is categorical data and independent variables are continuous data

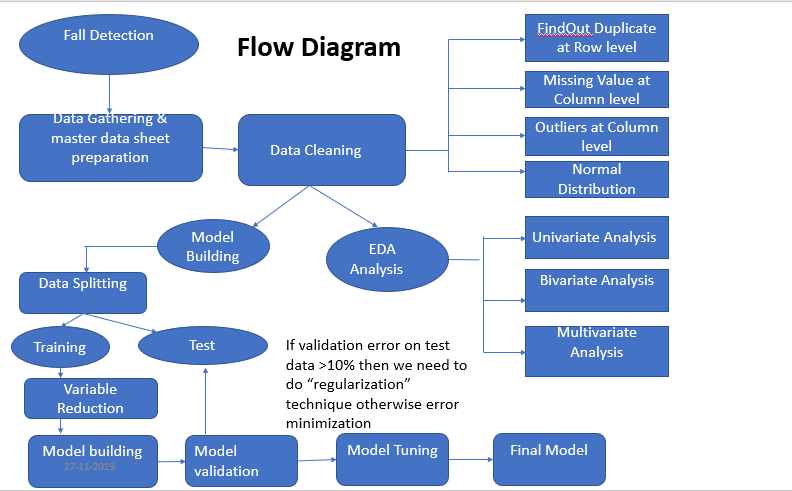
So, we are not able to find correlation between Dependent and Independent variables

1. Visual representation of all variables

# Approach

1. Load the data into Python, clean it and segregate the information into useful variables
2. Do EDA analysis find out useful information.
3. Split the Data set into Training and Test, build the model on Training Data set.
4. Do variable reduction technique to find best feature to build the model
5. Validate model on Test data sets and retune with help of deep learning algorithm to find the best accuracy of model and to avoid overfitting of model.

## 2.1 Flow Diagram



# Used Classification/Clustering algorithm

Classification is Process in which our dependent variable either Binomial or Polynomial in nature and All independent variable should be continuous in nature. Here our dependent variable has five categorical events, we must predict value 3🡪Falling with help of given continuous independent variable. We used logistic regression after changing ACTIVITY variable in from of 0 and 1 (Fall🡪1 and other then Fall🡪0) with Accuracy 0.7816235643409641

Clustering is a process of grouping a set of objects in such a way that the objects in the same group are homogenous and the objects in the different group are heterogeneous.

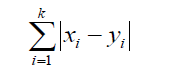
So, Clustering of ACTIVITY Colum with help of other given dependent variable we tried KNN clustering algorithm. This algorithm can be used if our variable either binomial our polynomial in nature. So, our data sets are full filling our criteria to use it.

This algorithm works on to find the distance of nearest neighbour with help of below methods

1. Euclidean Distance



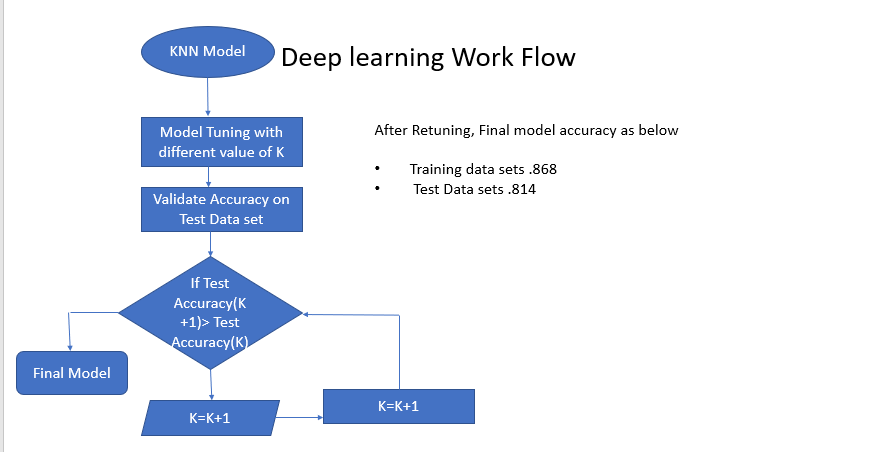
2-Manhattan Distance



With help of this algorithm we find our training and test accuracy as below

* Training Accuracy 🡪.7879
* Test Accuracy 🡪 .7845

To find the best accuracy we return the model and find value k=4 .



Final Accuracy after retuning model

* Training Accuracy 🡪.8679
* Test Accuracy 🡪 .8137

# Execution steps

## Step 1: Loading of data into Python

1. Load the csv data into Python
2. Once all the files are loaded, merge all the files into a single file

(In our case we have only file to upload, So we don’t have merging step)

## Step 2: Baseline model

1. Once the data is loaded, build a base line model for understanding the problem statement
2. At this step, we have used a small fraction of the actual data.
3. Remove Duplicates entry and find out the missing value
4. Treat outliers with help of IQR methods
5. Build the base model on logistic regression
6. Find proportional rational with dependent and independent variable
7. Find Accuracy of base model

## Step 3: Data pre-processing for final model

1. Load full data by merging all the available data into a single file
2. Pre-processing the file data cleaning like Remove duplicate entry, Checked missing value, treat outliers (We used IQR method to treat Outliers)
3. Activity column is having 6 categorical variables
4. Our Event rate is ~20%, So it will good to build the model
5. There are out layers in all Variable after 75% of records.
6. We have one duplicate entry in our data
7. Since we are using K mean Algorithm which is based on minimum distance major method, so to avoid influence of any variable we are using Minmax Algorithm method.
8. Find out best required variable to build the model, we used **variance inflation factor** (VIF) methods.
9. We used KNN Algorithm to build our model.

## Step 6: Retune the Model

1. We build our model on KNN Algorithm, For better accuracy we used different K value.
2. We got best value on training as well as test data on k = 4

# 5. Relationship With all Variables

1. we infer that before falling, the activity counts of Standing and Cramp are more

So, whenever we are getting activity code 0 or 4, we need to be alert.

1. As per mean value of all variables w.r.t Activity columns, we gather that Falling and Cramp have almost close values So, whenever we are getting activity Cramp, we need to be alert.
2. Event rate is ~21% (Total fall percentage in Activity Column)
3. We observed outliers in all variable as below.

**SL🡪** Observed Upper Outlaiers

**EEG🡪**Upper/Lower Outlaiers

**BP🡪** Observed Upper Outlaiers

**HR🡪** Observed Upper Outlaiers

**Circulation🡪** Observed Upper Outlaiers

# 6.Output and Recommendation

* If Sugar level increases, then chance of Falling will increase
* If BP level decreases, then chance of Falling will increase
* If EEG value will increase, then chance of Falling will increase
* If we get Cramp Activity code, we need to be alert
* We need to focus on three major Features SL, BP, EEG