Glossary

Obesity- Condition of being very fat
WHO- World Health Organization
Random Forest(RF)- Ensembled learning method
Logistic Regression- Model used for prediction and classification
Weka tool- GPU powered data pipeline
SMOTE- Used to reduce imbalance
SVM – Support Vector Mechanism
Hyperparameter Tuning- Parameter which increases learning process
Bias- inclination or prejudice for or against one person or group, especially in a way considered to be unfair
AdaBoostM2- statistical classification meta-algorithm
PCA- Principal Component Analysis
data normalization or standardization- Taking data into [0,1] or [-1,1] format
==

Data Preprocessing:

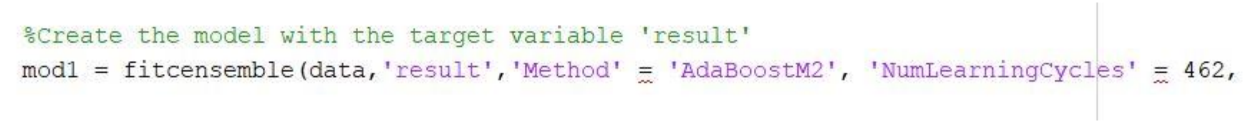
All of the data pre processing were done in python which is not mentioned in the poster. Starting from data cleaning to making it fit for model, etc. Besides basic EDA was also done in python.

Random Forest-

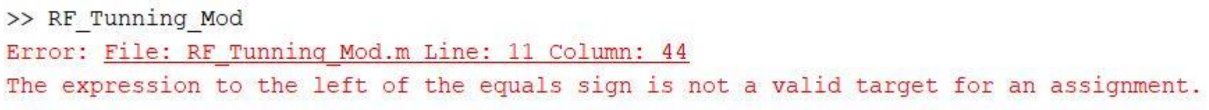
Hyper Parameter Tuning-
The iterated optimized parameter for our model are-

```
|======================================================================================================
==============================|
| Iter | Eval   | Objective   | Objective   | BestSoFar   | BestSoFar   |       Method |
NumLearningC-|    LearnRate |  MinLeafSize |
|      | result |             | runtime     | (observed)  | (estim.)    |              | ycles
|              |              |
|======================================================================================================
==============================|
|    1 | Best   |     0.22137 |      35.114 |     0.22137 |     0.22137 |          Bag |          319
|            - |          118 |
|    2 | Best   |     0.22041 |      22.103 |     0.22041 |     0.22089 |   AdaBoostM2 |          163
|    0.0057824 |            2 |
|    3 | Accept |     0.59176 |      57.152 |     0.22041 |     0.23097 |     RUSBoost |          427
|    0.0035486 |          403 |
|    4 | Best   |     0.17873 |     0.50334 |     0.17873 |     0.17888 |          Bag |           10
|            - |           46 |
|    5 | Best   |    0.080977 |      2.1604 |    0.080977 |    0.080981 |          Bag |           10
|            - |            1 |
|    6 | Accept |     0.83565 |       1.752 |    0.080977 |    0.081032 |   AdaBoostM2 |           13
|      0.62103 |         1035 |
|    7 | Accept |     0.83565 |      1.5241 |    0.080977 |    0.081118 |   AdaBoostM2 |           13
|     0.067994 |          967 |
|    8 | Accept |    0.60278 |    0.80646 |   0.080977 |   0.081021 |   AdaBoostM2 |            14
|    0.0012135 |           511 |
|    9 | Best   |   0.080498 |    0.52349 |   0.080498 |   0.080693 |          Bag |            10
|            - |             3 |
|   10 | Accept |    0.10541 |      2.496 |   0.080498 |   0.080659 |          Bag |            11
|            - |             9 |
|   11 | Accept |    0.24724 |     1.9441 |   0.080498 |   0.080553 |   AdaBoostM2 |            13
|    0.0012121 |             1 |
|   12 | Accept |   0.091998 |     2.0599 |   0.080498 |   0.080532 |   AdaBoostM2 |            15
|      0.96093 |             1 |
|   13 | Accept |    0.20364 |     1.6476 |   0.080498 |   0.080497 |   AdaBoostM2 |            14
|      0.18028 |             1 |
|   14 | Accept |    0.22712 |     4.3957 |   0.080498 |   0.080501 |     RUSBoost |            71
|      0.90558 |             1 |
|   15 | Accept |    0.86967 |     15.566 |   0.080498 |   0.080443 |     RUSBoost |            76
|      0.73723 |           914 |
|   16 | Accept |    0.23814 |     15.944 |   0.080498 |   0.080437 |     RUSBoost |            62
|     0.003504 |             1 |
|   17 | Best   |   0.077144 |     1.0315 |   0.077144 |   0.077455 |          Bag |            10
|            - |             2 |
|   18 | Best   |   0.071874 |     0.5243 |   0.071874 |   0.075412 |          Bag |            10
|            - |             2 |
|   19 | Accept |   0.077144 |    0.51759 |   0.071874 |   0.075849 |          Bag |            10
|            - |             2 |
|   20 | Accept |   0.074748 |     1.3885 |   0.071874 |   0.075596 |          Bag |            10
|            - |             2 |
|===============================================================================================
===============================|
| Iter | Eval   | Objective   | Objective   | BestSoFar  | BestSoFar  |       Method |
NumLearningC-|    LearnRate |  MinLeafSize |
|      | result |             | runtime     | (observed) | (estim.)   |              | ycles
|              |              |
|===============================================================================================
===============================|
|   21 | Accept |    0.12889 |     2.2381 |   0.071874 |   0.075383 |   AdaBoostM2 |            14
|      0.62177 |            22 |
|   22 | Accept |   0.095831 |     1.6334 |   0.071874 |   0.075325 |   AdaBoostM2 |            13
|       0.9862 |             6 |
|   23 | Accept |    0.24724 |     1.6138 |   0.071874 |    0.07534 |   AdaBoostM2 |            14
|    0.0011475 |            14 |
|   24 | Accept |    0.83565 |    0.36725 |   0.071874 |   0.075171 |          Bag |            10
|            - |          1006 |
|   25 | Best   |   0.054145 |     65.716 |   0.054145 |   0.054021 |          Bag |           494
|            - |             2 |
|   26 | Best   |   0.044082 |     83.024 |   0.044082 |   0.044178 |   AdaBoostM2 |           462
|      0.98188 |             2 |
|   27 | Accept |    0.24006 |     4.0532 |   0.044082 |    0.04417 |     RUSBoost |            37
|    0.0010315 |            10 |
|   28 | Accept |    0.22952 |     1.0197 |   0.044082 |   0.044164 |     RUSBoost |            13
|      0.83333 |             9 |
|   29 | Accept |    0.04552 |     54.197 |   0.044082 |   0.044064 |   AdaBoostM2 |           255
|      0.91604 |            10 |
|   30 | Accept |    0.23862 |     2.9829 |   0.044082 |   0.044061 |     RUSBoost |            12
|     0.047312 |             4 |

__________________________________________________________
Optimization completed.
MaxObjectiveEvaluations of 30 reached.
Total function evaluations: 30
Total elapsed time: 527.3336 seconds.
Total objective function evaluation time: 385.9994

Best observed feasible point:
      Method      NumLearningCycles    LearnRate    MinLeafSize
    __________    _________________    _________    ___________

    AdaBoostM2    462                  0.98188      2
```

This is the total iteration made inorder to find the optimal Hyperparameter for out model. This was not mentioned in the poster by me. Other iteration can be saught by us.
Because of using MATLAB2017, an old version I had to face some difficulties.
For example according to document I wrote the below code, which showed error



```
%Create the model with the target variable 'result'
mod1 = fitcensemble(data,'result','Method' = 'AdaBoostM2', 'NumLearningCycles' = 462,
```



```
>> RF_Tunning_Mod
Error: File: RF_Tunning_Mod.m Line: 11 Column: 44
The expression to the left of the equals sign is not a valid target for an assignment.
```

I had to resolve this by understanding the code in my version of software

Logistic Regression

I was getting error and still I was trying with glmfit() function in my MATLAB file, I resolved it using mnrfit() function after understanding one is for linear regression and the other is for multinomial class regression

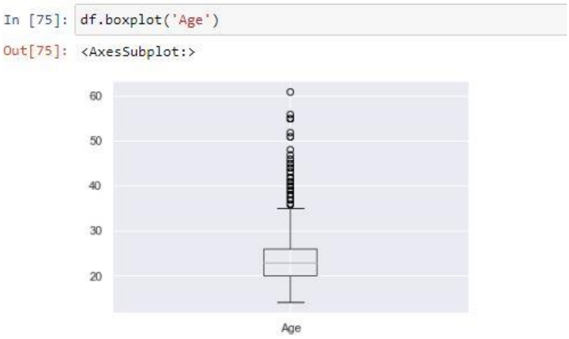
Other Tried method, worth to mention

For Random Forest, I did 5 fold cross validation, before doing that I tried using CV partition and splitting the datasets into equal parts of 80 and 20 percent,
The accuracy differed by some small margin in that case-

Accuracy with 5- Fold CV- 95.5%
Accuracy without 5- Fold CV- 95.3%
The difference is very low

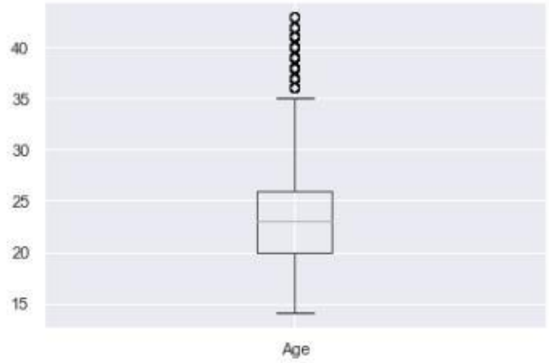
Python Code

Here are some mentionable python works which I did-



```
In [75]: df.boxplot('Age')
Out[75]: <AxesSubplot:>
```

This was the number of outliers before I removed it, after I removed the outliers of age column, the result became not totally clean but was looking better-



Besides this label encoding using ML libraries of python and creation of heatmap were some of the most important works I did in python.