First of all I have put my requirements libraries in the submission For this task since I had imported sklearn,tensorflow, numpy,pandas, matplotlib so you need to have this libraries first

In task 2 first phase is the data preprocessing

1 - First thing I have noticed that variable 2 in composed of 2 variables so I split the variable into 2 variables

2 – Variable 18 have a lot of nan (More than 50%) So I decide to remove it from the dataset .

3 – Variable 19 Is overfiting the training set since when classlabel = no Variable 19 = 0 And if classLabel = yes Variable 19 = 1 But this is not happening in the validation set So I decide to drop this column .

4 - variable5 is redundant out of variable4 since their values is in same pattern so I decide to remove variable 5

5 - there is a relation between var10 and var11 as if var10 = f then var11 = 0 else var 10 will be numerical number

6 - variable 17 = variable 14 \*1000 so variable 17 is redundant variable "Droped it "

7 – all variables with nan on it if It is char then I will replace it with the most used char else I will used the mean except for variable 2-1 , variable2.2 I replaced it with the median since they have a finite possible numbers

8 – replacing all the char with number instead

9 – I notice also in data modeling phase that variable 9 is one of the most effective variable in dataset

After this I choose Logistic regression algorithm for this problem to classify (Yes or No ) With normalization I got **0.72** accuracy in validation set without normalization I got **0.68**

Then I tired to use support vector machine but I didn't get a good accuracy so I removed it .

Then I Used deep neural network for this problem with 7 layers Neural in each layer (128,,64,32,8,4,2,1) With **relu** for all as activation function except for last layer (output layer) I used activation function **sigmoid** With layers and with normalization the data I got **0.80** accuracy in validation data and there is was an overfiting on the data too After that I used regularization like L2 and dropout but my accuracy of validation set didn't go well so I decide to remove it

As a conclusion I Got max accuracy 0.8 on validation set using DNN with normalized data