1. Install Matlab
2. Download the code ‘gmm.m’
3. Download the test file in *csv* format.
4. Import the *test.csv* in matrix format and generate a script
5. Write a command   
    **data=test**  
   below the script
6. Copy the contents of code ’gmm.m’ below the above mentioned script.
7. Save the above script and run it.
8. You will get two classes of answer.
9. The class of answer which you want will be clear when you will do the same procedure to the *train.csv* without the target and by comparing the target with the two classes of answer.

**NOTE: In the 5th step data should be equal to the filename being imported.**

The data contained both the numeric and categorical data. To incorporate the categorical data I assigned numeric values to particular alphabets but it didn’t work out because the data also contains the values like “XYZ”. So I dropped the idea of incorporating categorical data as there were time constraints also. As it was a classification problem I learned the machine learning algorithm for classification problems which can also gave me the probability. I learned the **clustering – Gaussian mixture models** and coded them. After removing the categorical columns manually i imported the data ’*test*.csv’ in matrix form in matlab and run it. The code removes the rows of data containing empty values and hence the probabilities of the particular ID’s cannot be found.