**Project Report**

**1-training data**

**2-preprocessing**

**3-classification**

**3-classification:**

Input:

1-3 text files (“easy\_ham” , “hard\_ham”,”spam”) containing a counter (data structure: a list of tuples ) the counter has a number of words that were output of the preprocessing stage linked with the number of their ocurrence in each case of easy ham ( easily identified non-spam mail) , hard ham (hardly identified non-spam mail) and spam.

2-an email to classify input into file

**Output:**

Command line : “spam mail” or “non-spam mail”

**Process:**

1-changing text files manually into .py extension and making three variables spam,eham,hham equal to lists

On importing these files we can use lists directly from any file