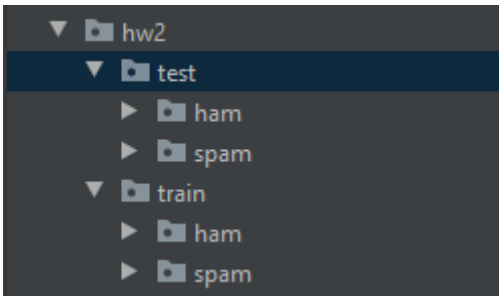
**Steps to execute the code:**

- Copy the complete folder Naive_Bayes_Logistic_Regressions to a particular location.
- Execute the code in Linux System- else you might have to change "//" to "\\" in read_convert_data function in convert_data_to_dataframe.py file.
- The linux version should be 3 or more should work good for 2.7. Numpy,pandas,sklearn,regex dependency required.
- The "programs" folder has the complete code.
- Following are the program names to execute :
- The input folder structure should be as in the screen shots below:




Need to give path till "\hw2\train" or "\hw2\test"

- generate_vectors.py for generating both Bernoulli and Bag_of words vectors. Once this executed six excel files will get generated in programs folders. These are train and test vector files for each of both Bernoulli and bag_of_words.
- **Every time you execute a new dataset. Do re-execute generate_vectors.py before executing other files as it generates new vector file specific to the dataset which is supposed to be used by other algorithms.**
- Execute as following below
  python generate_vectors.py <train_location> <test_location>
  python generate_vectors.py "D:\python generate_vectors.py "D:\Naive_Bayes_Logistic_Regressions\hw2\train" "D:\Naive_Bayes_Logistic_Regressions\hw2\test"\hw2\train" "D:\Naive_Bayes_Logistic_Regressions\hw2\test"

python naive_bayes.py <bag_of_words or bernolli> please give exact spellings. "python naive_bayes.py bag_of_words " generates results for multinomial naive bayes with bag_of_words and "python naive_bayes.py bernolli " generates results for Bernoulli vectors with discrete naive bayes.

python  logistic regression.py <bag_of_words or bernolli> for logistic regression.

python sgd_classifier.py <bag_of_words or bernolli> for sgd