When the classifier was run for the given dataset with 3183 examples in six categories we tested the given classifier on various values of K. With each increase in value of K, We were able to record percentage accuracy for that for that value of K. Graph1 depicts clearly the relation between the increasing value of K and its effect on the classification of document different categories. It is evident that the classifier was able to achieve accuracy of greater than 94% for K = 5, i.e. when four fold of the data was used as training set to build the classifier and rest for testing the classifier. We used the various methods of accuracy measurements like recall, Precision, and F-measure to verify the accuracy of our classification approach. Using our approach we were able to generate quite a large

Table IV shows the results obtained when four folds of 3183 examples were used as the training set to build the

Classifier and the remaining fold 489 examples were used to test the classifier for accuracy. We use recall [25], precision [25] and F-measure [25] to verify the accuracy of our classification approach. F-measure is the harmonic mean of recall and precision. Recall, Precision and F-Measure are calculated as follows:

Using the three measures, we observe that the average precision is 89.09%, average recall is 89.04%, whereas the

F-Measure is 89.05%. Thus, classification of web sites is possible by examining the contents of their home pages.

*Number of Training Examples and Accuracy*

The classifier was subjected to training and testing in 9 steps each time increasing the input by 50 documents.

Graph 1 depicts the number of training examples versus the accuracy in terms of average F-measure. The accuracy of the classifier was very poor i.e., about 45%, when only 50 documents were supplied as training data. The accuracy increases each time when the classifier is supplied with additional learning data. The classifier achieved an accuracy of 89% when nearly 450 documents were supplied as input in each category.

Thus, the accuracy of the classifier depends on the number of training documents and in order to achieve high accuracy, the classifier should be supplied with sufficiently large training documents.