**TESTING**

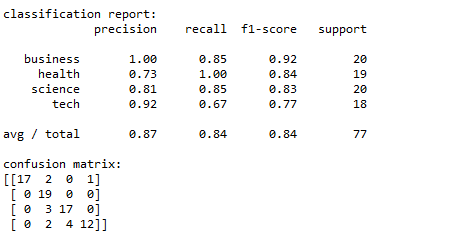
**Test Cases:**

1. **Size of Dataset:**

In machine learning application size of dataset is very important**.** We have to make sure that the because of less no of data the algorithm might under-fit the testing data. And if we have large amount of data then algorithm may over-fit the testing data.

* Classification with less training dataset :

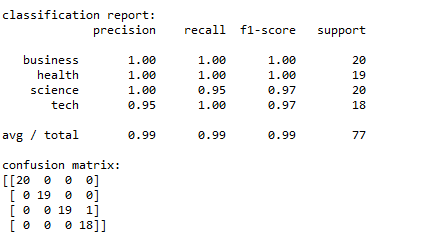
The following is the classification report after running the Linear SVM on test dataset.



As we can see there are some classification of news going wrong. The F1-score is about 0.84 which is not that good

* Classification with appropriate training dataset:

The following is the classification report after running the Linear SVM on test dataset.



As we can see only 1 news is getting classified as wrong. The F1 score is 0.99 which very good.

Thus we can see that appropriate dataset can improve the algorithm.

1. **Number of news per story for clustering:**

The number of news per story can range from 1 to 5[max].Since occurrence of news of particular story at different sources occurs at different time. So the news should be clustered accordingly.

* Story with 1 news should be clustered differently from other news. The following last news about “Asus Fonepad 7 phablet for Rs 12,999 ”
* Story with 4 news should not include any other 5th news. The following news about “Nokia X smartphone ” contains only 4 news

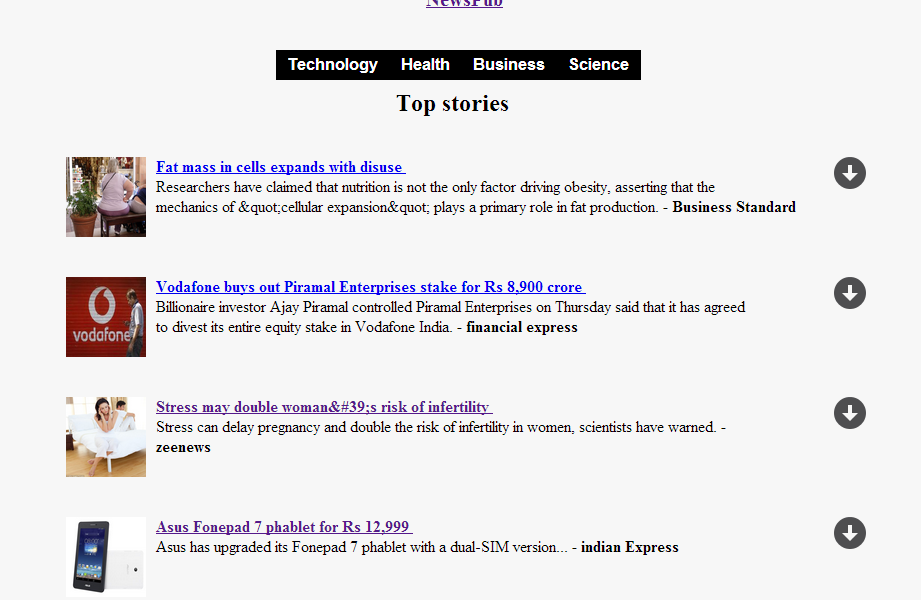


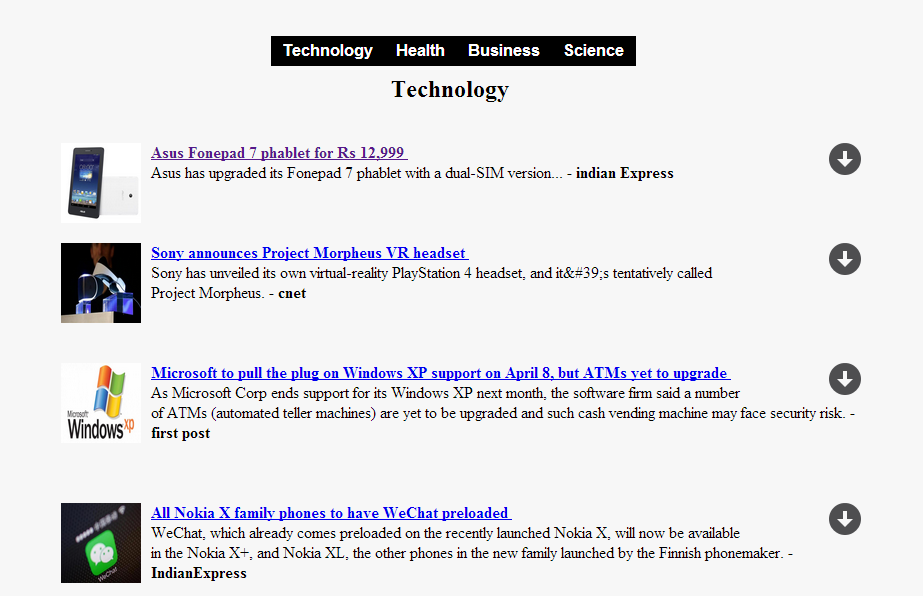
* The below figure shows clustering of 2 news and 5 news.



1. **Clustering of the mixed news i.e of all the category [health, science, business, Technology] together and clustering of news in a specific category eg. Technology**

The following figures shows mixed news cluster and specific category news respectively.





1. Html and css working properly

**CONCLUSION:**

* Size of the training dataset have effects on accuracy of the algorithm. The small size of training dataset results in some news wrongly classified leading to decrease in accuracy of classification algorithm.
* When the classification algorithm is provided with appropriate size of dataset it results in very high accuracy and hence we observed that only 1 news was wrongly classified
* Proper clustering of news story with different number of occurrence 1,2,3,4… is an important aspect of the application. As we have observed that news with only 1 occurrences is clustered different from other news.
* We have also observed that other news story with different no of occurrences is clustered accordingly. The different news stories are not mixed up together to form a cluster.
* We have observed that when all the stories of different categories are mixed for the main page of application then the news are clustered well and no problem has occurred
* Also when news of a category are clustered then the news are clustered well and no problem has occurred. There are no ambiguity in the clustering the news.