

WordCloud

Για το wordcloud χρησιμοποιήσαμε τις βιβλιοθήκες απο:

https://github.com/amueller/word_cloud

και συμβουλευτήκαμε τα tutorials από εκεί για να κατασκευάσουμε το wordcloud μας

Clustering

Για το clustering χρησιμοποιήσαμε την βιβλιοθήκη :

Nltk.sklearn ώστε να πάρουμε από εκεί τη συνάρτηση Kmeans και ο,τι χρειαστήκαμε για την cosine distance. Επίσης συμβουλευτήκαμε το site

<https://docs.python.org/2/tutorial/classes.html> και ορισμένα video από το youtube ως tutorial για την κατασκευή της κλάσης που χρησιμοποιήσαμε.

Clusters	Technology	Politics	Film	Business	Football
1	0.01	0.1	0.0	0.89	0.0
2	0.68	0.02	0.01	0.21	0.08
3	0.0	0.97	0.0	0.02	0.0
4	0.0	0.0	0.0	0.0	0.99
5	0.02	0.01	0.96	0.0	0.0

Classification

Από τους classifiers κάναμε τον Bernoulli Naive-Bayes , SVM και Random Forests.

Χρησιμοποιήσαμε τις βιβλιοθήκες που παρέχει το scikit learn. Κάναμε το 10-fold Cross Validation χρησιμοποιώντας τις μετρικές Accuracy , Precision , Recall και F measure χρησιμοποιώντας επίσης τις βιβλιοθήκες του scikit learn.

Statistic Measure	Naïve Bayes	Random Forest	SVM	KNN
-------------------	-------------	---------------	-----	-----

Precision	0.933391309792	0.910379869824	0.919842459219	0
Accuracy	0.935024616068	0.910484929223	0.906976724089	0
Recall	0.927585073735	0.893717270419	0.878219654886	0
F_Measure	0.929923020186	0.899308951019	0.887535550221	0

Για να τεστάρουμε τα test_data που μας δώσατε χρησιμοποιήσαμε τον classifier Naïve-Bayes.

ID	Predictions
2	Politics
10	Technology
25	Films

·
·
·

15332	Football
15333	Football