**Softwares used:**

**Spyder 3** (Scientific PYthon Development EnviRonment)

**Language used:**

**Python 3.5**

**Libraries used:**

* **pandas** (Python Data Analysis Library) for inputting the Dataset (.tsv file).
* The module **pyplot** of **matplotlib** library is used for graph plotting
* **re** library for the usage of Regular Expressions for text cleaning.
* **Stopwords** list from the ‘**corpus’** package of **nltk** (Natural Language Processing Toolkit) data package.
* **PorterStemmer** algorithm from the package **nltk.stem.porter** for stemming of words.
* **CountVectorizer** class from **sklearn.feature\_extraction.text** submodule for building the matrix (sparse) of word counts from text documents.
* **TfidfTransformer** class from **sklearn.feature\_extraction.text** submodule to convert the count matrix (sparse) to a matrix of TF\_IDF features.
* **TruncatedSVD** class from **sklearn.decomposition** module for dimensionality reduction of the feature space.
* **train\_test\_split** function from **sklearn.cross\_validation** module for splitting the whole Dataset into random train and test sunsets.
* **SVC** class from **sklearn.svm** module for classification of the test set.