

AKSHAT ANAND
14055
Assignment 2
CS671 (NLP)

The submission contains 23 files in total. 20 of those are codes containing feature extraction and classifier. 3 are for preprocessing the text, creating the word2vec model and creating doc2vec model. For, glove I have taken the glove pre-trained word vector from [Stanford \(Wikipedia 2014 + Gigawords\)](#). I have used 50d (50 dimension) file from that. So, first run word2vec.py and doc2vec.py and then continue with the original codes. Extract complete glove zip file in the folder where codes are. So, it will create the desired directory structure. Now, the 20 codes are described below.

Feature Extraction	Classifier	Python Code	Accuracy
Bag of Words	Logistic regression	BoWLR.py	88.58
Bag of Words	Naive Bayes	BoWNB.py	83.42
Bag of Words	SVM	BoWSVC.py	83.40
Normalized tf	Logistic regression	tfLR.py	86.56
Normalized tf	Naive Bayes	tfNB.py	83.12
Normalized tf	SVM	tfSVC.py	64.34
Tfidf	Logistic regression	TfidfLR.py	88.76
Tfidf	Naive Bayes	TfidfNB.py	83.36
Tfidf	SVM	TfidfSVC.py	76.19
Word2Vec	Logistic regression	W2vLR.py	83.68
Word2Vec	SVM	W2vSVM.py	82.97
Word2Vec	Feed Forward (Vanilla)	W2vFeed.py	84.10
Word2Vec	LSTM	W2vLSTM.py	83.60
Glove	Logistic regression	GloveLR.py	76.24
Glove	SVM	GloveSVC.py	75.72
Glove	Feed Forward (Vanilla)	GloveFeed.py	79.05
Glove	LSTM	GloveLSTM.py	78.30
Doc2Vec	Logistic regression	d2vlr.py	77.54
Doc2Vec	Feed Forward (Vanilla)	D2vVanilla.py	76.08
Doc2Vec	LSTM	D2vLSTM.py	76.42