**Kshitiz Suman 14333**

I have used Binary Bag of Words using sklearn library's method CountVectorizer.

I have used Normalized Term Frequency and Tfidf Representation using TfidfTermVectorizer.

For Word2vec representation I have used Google PreTrained Word2vec representation.

For Glove representation I have used Stanford PreTrained Representation.

In case of Word2Vec and Glove ,I have used GaussianNB in Naive Bayes for Sentiment Analysis because these type of representation contains negative elements and MultinomialNB will give an error.

In case of Bag Of Words, Tf, Tfidf I have used MultinomialNB in Naive Bayes for Sentiment Analysis.

**Logistic Regression** **Naive Bayes**  **SVM Classifier**

(MultinomialNB/GaussianNB)

**Binary Bag of Words**  85.924 81.98 63.092

**Normalized Term Frequency**  79.588 84.688 60.43

**Tfidf Representation**  88.032 83.028 67.8

**Word 2 Vec** 85.264 76.472 79.02

**Word 2 Vec with Tfidf**  78.056 66.996 50.192

**Glove Representation** 79.679 71.064 77.8

**Glove with Tfidf** 74.204 61.148 64.259

From the above data trend, we can see that there is a very high dependency on the type of representation chosen and the algorithm used . An algorithm performing well for one representation may or may not perform well for the other representation . For ex :- The Word2Vec representation used on SVM Classifier gives 79% accuracy on test data while Word2Vec with Tfidf on SVM Classfier gives only 50.192 percent .Overall the Tfidf Representation with Logistic Regression shows the best result of 88.032% as compared to other results.There is a general trend for the Word2Vec representation with Tfidf and without Tfidf.Without Tfidf the accuracy is more .Naive Bayes performs better than Logistic Regression in case of Normlized Term Frequency while it fails for all the other data type representaion.