

# **Assignment 3**

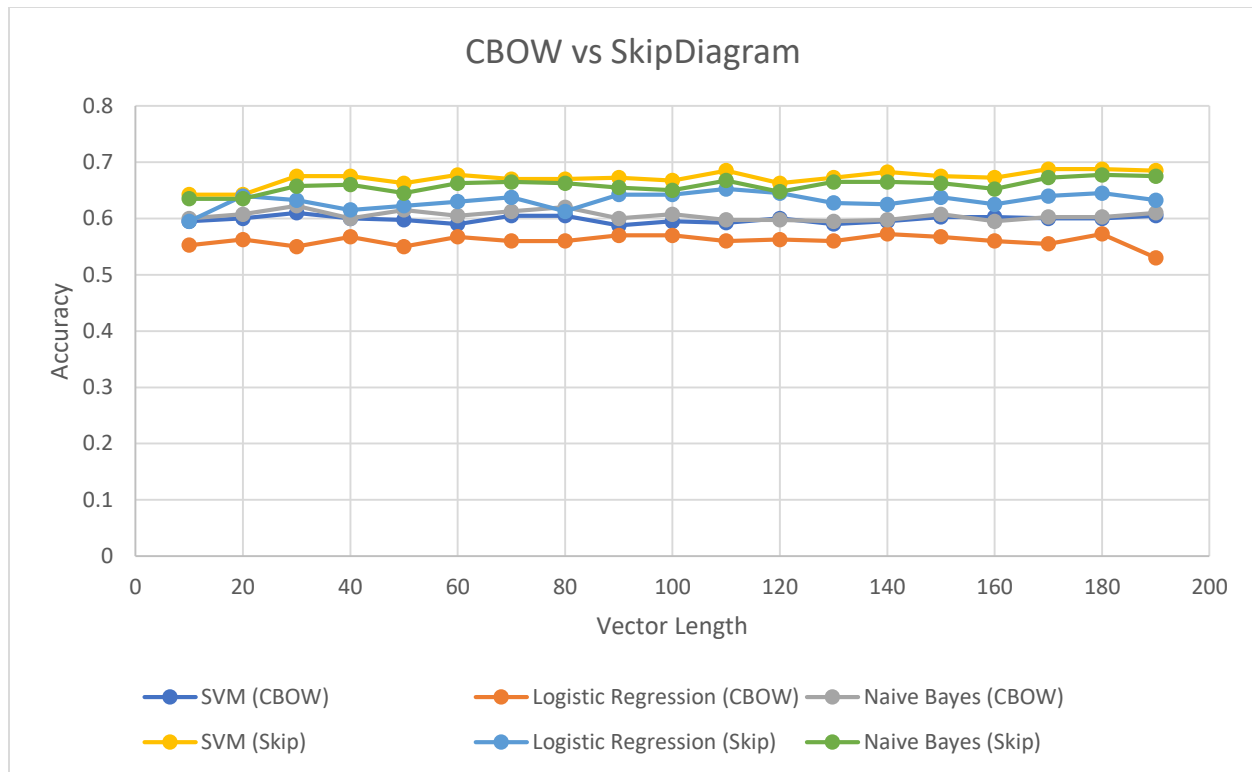
**CS462**

**NLP**

Hatem Mamdoh Saeed	20170085
Hussien Ashraf Feteiha	20170093

	<b>CBOW</b>	<b>Skip Diagram</b>	<b>TF-IDF</b>
<i>SVM</i>	0.61	0.6875	0.8675

	<i>CBOW</i>			<i>Skip Diagram</i>		
<i>Vector Length</i>	<i>SVM (CBOW)</i>	<i>Logistic Regression (CBOW)</i>	<i>Naive Bayes (CBOW)</i>	<i>SVM (Skip)</i>	<i>Logistic Regression (Skip)</i>	<i>Naive Bayes (Skip)</i>
10	0.595	0.5525	0.6	0.6425	0.595	0.635
20	0.6	0.5625	0.6075	0.6425	0.64	0.635
30	0.61	0.55	0.6225	0.675	0.6325	0.6575
40	0.6	0.5675	0.6	0.675	0.615	0.66
50	0.5975	0.55	0.615	0.6625	0.6225	0.645
60	0.59	0.5675	0.605	0.6775	0.63	0.6625
70	0.605	0.56	0.6125	0.67	0.6375	0.665
80	0.605	0.56	0.62	0.67	0.6125	0.6625
90	0.5875	0.57	0.6	0.6725	0.6425	0.655
100	0.595	0.57	0.6075	0.6675	0.6425	0.65
110	0.5925	0.56	0.5975	0.685	0.6525	0.6675
120	0.6	0.5625	0.5975	0.6625	0.645	0.6475
130	0.59	0.56	0.595	0.6725	0.6275	0.665
140	0.595	0.5725	0.5975	0.6825	0.625	0.665
150	0.6025	0.5675	0.6075	0.675	0.6375	0.6625
160	0.6025	0.56	0.595	0.6725	0.625	0.6525
170	0.6	0.555	0.6025	0.6875	0.64	0.6725
180	0.6	0.5725	0.6025	0.6875	0.645	0.6775
190	0.605	0.53	0.61	0.685	0.6325	0.675



All the above tables were run on a train\_test\_split random state = '3'

It can be seen that on this corpus, the tf-idf performs better than both the CBOW and skip diagram scoring higher accuracy. Also it can be seen that the skip diagram performs higher than the CBOW.

From the graph:

- The best vector length for the SVM (CBOW) was 30.
- The best vector length for the Logistic Regression (CBOW) was 140.
- The best vector length for the Naïve Bayes (CBOW) was 30.
- The best vector length for the SVM (Skip Diagram) was 170.
- The best vector length for the Logistic Regression (Skip Diagram) was 110.
- The best vector length for the Naïve Bayes (Skip Diagram) was 180.