|  |  |  |
| --- | --- | --- |
| الحسن ابراهيم جادالرب | 20191700125 | SC/ sec:1 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Preprocessing**:

* Bayse & tf-idf
  + Stemming
  + Stop word removal
  + Punctuation removal
  + Word tokenization
* Word2Vec
  + Stop word removal
  + Punctuation removal
  + Word tokenization

**Feature extraction:**

* Bayes
  + Word positive to negative ratio
* Tf-idf
  + Tf-idf: term frequency multiplied by one over term document frequency
* Word2vec
  + Word embeddings

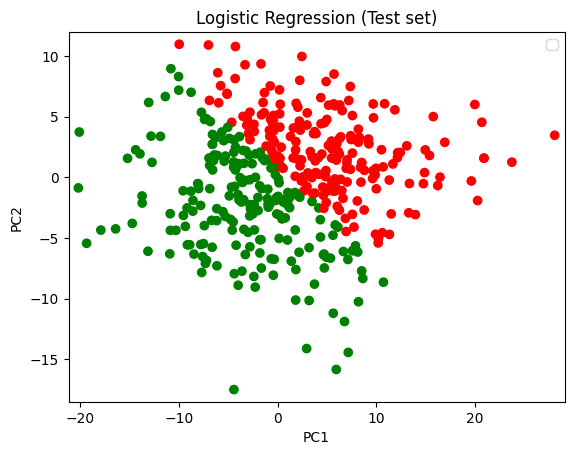
**Classification:**

* **Logistic regression**
  + **Tf-idf 🡺Train: 97% , test:84%**
  + **Word2vec 🡺 train :81%, test:79%**
* **SVM**
  + **Tf-idf 🡺Train: 99% , test:86%**
  + **Word2vec 🡺 train :79%, test:78%**
* **Bayes** 
  + **Accuracy : 75%**

**TF-IDF**

**A picture containing text, screenshot, diagram, plot

Description automatically generated**

**WORD2VEX**