Lab Exercise 9 - Word Embeddings

In this lab exercise, we will create word vectors (embeddings) using word embedding algorithms. Apply these algorithms for the twitter data (extracted in the earlier exercise) or on any corpus of your choice (https://www.corpusdata.org/formats.asp).

1. Apply the following word embeddings on the concerned corpus:
    a. GloVe
    b. Word2Vec
    c. FastText

You may update the python notebooks shared.

Check the correctness of the model by plugging in word similarities as shown in Figure 1.

```
words = ['draupadi','krishna','king','arjuna','pandavas','kunti','duryodhana']
for i in range(len(words)):
    print(words[i], end="\t==>  ")
    similar = cbow.wv.most_similar(words[i], topn = 5)
    for j in range(len(similar)):
        print(similar[j][0],end =", ")
    print("\n")
```

```
draupadi        ==>  drupada, madri, virata, subhadra, duhsasana,

krishna ==>  kesava, vasudeva, govinda, janardana, rama,

king    ==>  monarch, sire, janardana, bharata, madhava,

arjuna  ==>  karna, dhananjaya, bhima, partha, bhimasena,

pandavas        ==>  kauravas, parthas, dhartarashtras, kurus, pancalas,

kunti   ==>  pritha, pandu, radha, dharma, subala,

duryodhana      ==>  karna, dhritarashtra, jayadratha, suyodhana, bhima,
```

Fig. 1 : Word similarities on Mahabharatha Corpus.