

# Assignment Set 11

Pegah Khazaei

24<sup>th</sup> December 2020

---

When the text has been provided, the computer will utilize algorithms to extract meaning associated with every sentence and collect the essential data from them.

The first step in this regard is to convert the sentences in our corpus into tokens or individual words.

The next step is to create a dictionary that contains all the words in our corpus as keys and the frequency of the occurrence of the words as values.

Use of dictionary helps us convert unstructured text into structured data. The more precise dictionary you have for the analysis, the more accurate will be the analysis or prediction.

Word embeddings provide a dense representation of words and their relative meanings. They are an improvement over sparse representations used in simpler bag of word model representations. Word embeddings can be learned from text data and reused among projects.

