Natural language processing is employed in this endeavour to educate computers how to comprehend human languages (NLP).

To complete this task, we used a number of nlp techniques.

After starting with some plain text, a nlp model goes through a lot of phases.

The tokenizer separates the input's raw form into individual words.

Stopwords are used to eliminate all the superfluous and frequent words once the lemmatizer reduces these terms to their most basic forms.

When all processing is finished, the clean text is what is left, and the model is given this data.

Since computers can only interpret numbers, the process then begins by turning this cleansed data into numbers.

This procedure  can carried out differently

A few nlp approaches, such as bag of words, tf-idf, and word embedding, enable us to represent text data into numerical form.

In this project, TF-IDF will be used.

In this method, the priority of the word is determined by counting the number of times each character appears and dividing it by the total number of words which called the TF of the word , then we take log of number of numbers of words divided by