Assignment - NLP

1.	Consider the text of following documents: [CBSE, 2022] Document 1: Sahil likes to play cricket Document 2: Sajal likes cricket too Document 3: Sajal also likes to play basketball Apply all the four steps of Bag of words model of NLP on the above given documents and generate the output.
2	Through a step-by-step process, calculate TFIDF for the given corpus: **Document 1:** Johny Johny, Yes Papa, **Document 2:** Eating sugar? No Papa **Document 3:** Telling lies? No Papa **Document 4:** Open your mouth, Ha! Ha! Ha!
3	Following is the corpus: **Document 1:** This is my first experience of Text Mining. **Document 2:** I have learnt new techniques in this. Perform the following functions on the above corpus: a. Sentence Segmentation b. Tokenization c. Stop words removal d. Lowercase conversion
4	Create a document vector table from the following documents by implementing all the four steps of Bag of words model. Also depict the outcome of each step. Document 1: Sameera and Sanya are classmates. Document 2: Sameera likes dancing but Sanya loves to study mathematics. 4
5	Normalize the given text and create a document vector table of the corpus of the following text: *Harry and Garry play cricket every day. They also drink milkshakes every day. Harry loves chocolate milk and Garry loves strawberry milkshakes.*
6.	Following is the corpus: **Document 1:** This is my first experience of Text Mining. **Document 2:** I have learnt new techniques in this. Perform the following functions on the above corpus: a. Sentence Segmentation b. Tokenization c. Stop words removal d. Lowercase conversion