**WORKSHEET-1**

**NLP**

1. Which of the following are steps in NLP?

Answer: A) Lexical Processing B) Syntactic processing C) Vectorizer processing

2. Which of the following tasks can be completed with only lexical processing?

Answer: B) Machine Translation D) Question- Answering system

3. Which of the following are steps in lexical processing?

Answer: C) Top-Down parsing D) POS-tagging

4. Which of the following tokenizers are available in NLTK?

Answer: A) word\_tokenize() B) sent\_tokenize()

5. What will be the output of the following lines of code?

Answer: A) [“I”, “love”, “#”, “food”, “#”, “pasta”]

6. What will be the output of the following lines of code?

Answer: B) [“I”, “love”, “#food”, “#pasta”]

7. Which of the following is/ are true regarding to stopwords?

Answer: C) removing stopwords results in faster computation.

8. In which of the NLP tasks we can remove stopwords?

Answer: A) spam-ham classifier building

9. Which of the following is/are true regarding bag of words model of text?

Answer: A) It takes in to consideration of only the words present in the text and not the order of the words.

C) It captures the semantics of the text.

10. Consider the following two documents we create a bow representation using Count Vectorizer of NLTK library. What will the shape of the resultant data?

Answer: B) (2,5)

11. Which of the following are true regarding Tf-Idf?

Answer: C) All the words are treated equally regardless of whether they are present in other documents or not

12. What will be the tf-idf score of word “Bangalore” in Document 1?

Answer: A) 0.2

13. What will be the tf-idf score of the word “beer” in document 2?

Answer: B) 0.89

14. Which of the following statements are true regarding the above documents?

Answer: B) The tf-idf score of “vapour” is less than tf-idf score of “Bangalore” in document 1

15. Which of the following are advantages of using tf-idf model over BOW model?

Answer: