PAPER ID—20587

P.G. Diploma in Business Analytics

EXAMINATION, 2023

(Second Semester)

NATURAL LANGUAGE PROCESSING

Time : 3 Hours   *Maximum Marks : 35*

Before answering the question-paper candidates should ensure that they have been supplied to correct and complete question-paper. No complaint, in this regard, will be entertained after the examination.

Note : There are two Sections A and Section B. Section A has seven questions, each carries 1 marks. Section B has *four* questions, each carries 7 marks. In Section B, students must choose *one* of the two questions (*one* from each Unit).

Section A

1. What is Spacy ?   1

2. Explain cluster analysis.   1

3. What is Gensim ?   1

4. What does NLP stand for ?   1

5. Explain the concept of stemming.   1

6. What are the stopwords ?   1

7. What is the role of POS tagging ?   1

Section B

Unit I

1. What is a stemming algorithm and its different types ? Explain by giving an example using Python codes.   7

Or

Explain the concept of tf-idf's law in NLP by using example data.

Unit II

2. What is cluster analysis and what are its different categories ?   7

Or

Explain the process of K-means Text Clustering using example.

Unit III

3. Explain the role of lemmatization using Python codes.   7

Or

What are the advantages and disadvantages of using n-grams for sentiment analysis ?

Unit IV

4. Explain the topic model using LDA.   7

Or

What is the use and role of HMM ?