

Exam No: _____

GANPAT UNIVERSITY

B. TECH SEM-VI (COMPUTER ENGINEERING (ARTIFICIAL INTELLIGENCE))

FIRST INTERNAL EXAMINATION – FEBRUARY-MARCH 2023

2CEAI602: Natural Language Processing

TIME: 1 Hour

TOTAL MARKS: 20

- Instructions:**
- 1) Figures to the right indicate full marks.
 - 2) Be precise and to the point in your answer.
 - 3) The text just below marks indicates the Course Outcomes Numbers, (CO) followed by the Bloom's taxonomy level of the question, i.e., R: Remembering, U: Understanding, A: Applying, N: Analyzing, E: Evaluating, C: Creating.

- Q.1** How the ambiguity, common knowledge, and human creativity pose a challenge in the context of Natural Language Processing? **[4]**
1U
- Q.2** Consider the input dataset as given below in variable data. Write a python code to read the data and apply data cleaning techniques as stop words removal and lemmatization. Apply CountVectorizer on a cleaned dataset to generate output matrix. **[4]**
3A
data = ["We are GUNI students",
"At GUNI campus, we are learning Artificial Intelligence",
"GUNI is having large green campus"]
- Q.3** Elaborate the various phases of the Skip Gram model that are utilized in Word2Vec with appropriate diagram? **[5]**
2R
- Q.4** Describe TF-IDF briefly. Apply TF-IDF for following Corpus: **[7]**
1A
D1: "The Health Observations for March"
D2 "The Health oriented calendar"
D3: "the Awareness News for march Awareness"
Query: "March health awareness"
Apply above query to find similarity using Cosine function and magnitude.

-----END OF PAPER-----