Name: Enrolment No:	

UPES

End Semester Examination, December 2023

Program Name: MCA **Semester : 3rd**

Course Name : Natural Language Processing **Time : 3 hr**

Course Code : CSAI8004P **Max. Marks : 100**

Instructions :

1. **Attempt all the sections.**
2. **The question paper consists of 11 questions, and it is divided into three sections A, B, and C.**
3. **There is an option for question no. 8. Solve anyone.**
4. **There is an option for question no. 11. Solve anyone.**

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Define: • Precision • Recall • Cosine similarity • Perplexity	4	CO1
Q 2	Write a difference between the intrinsic and extrinsic evaluation.	4	CO1
Q 3	Explain the four basic applications of NLP.	4	CO2
Q 4	Describe the Dependency Parsing.	4	CO1
Q 5	Define the term frequency-inverse document frequency (tf-idf).	4	CO2
SECTION B **(4Qx10M= 40 Marks)**			
Q 6	Discuss the components of a typical question answering system and their roles in the system.	10	CO2
Q 7	Discuss the typical preprocessing steps involved in sentiment analysis, such as text tokenization, stop word removal, and stemming/lemmatization. Also, elaborate the important of these steps and their contribution in the accuracy of sentiment analysis?	10	CO1
Q 8	Explain the different types of questions processed by the question answering system. OR Explain the steps of NLP and their significance with one of the NLP applications.	10	CO5

<table>
<tr><td>Q 9</td><td>Calculate the coefficient of correlation of the following data by Spearman's rank correlation method:

<table><tr><td>X</td><td>19</td><td>24</td><td>12</td><td>23</td><td>19</td><td>6</td></tr><tr><td>Y</td><td>9</td><td>22</td><td>20</td><td>14</td><td>22</td><td>18</td></tr></table></td><td>10</td><td>CO5</td></tr>
<tr><td colspan="4">SECTION-C
(2Qx20M=40 Marks)</td></tr>
<tr><td>Q 10</td><td>In the context of the Vector Space Information Retrieval (IR) model, you are provided with 5 documents and 2 queries. Your goal is to use this model to retrieve the most relevant documents for each query.

Documents:
Document A: "The quick brown fox jumps over the lazy dog."
Document B: "A watched pot never boils."
Document C: "To be or not to be, that is the question."
Document D: "In the depth of winter, I finally learned that within me there lay an invincible summer."
Document E: "Two roads diverged in a wood, and I—I took the one less traveled by."

Queries:
Query 1: "What is the meaning of life?"
Query 2: "How can I stay motivated during difficult times?"</td><td>20</td><td>CO4</td></tr>
<tr><td>Q 11</td><td>Consider the following context-free grammar and input string for a shift-reduce parser task:

Grammar:
S → S+S
S → S-S
S → (S)
S → a

Input string:
a1-(a2+a3)</td><td>20</td><td>CO3</td></tr>
</table>

	Please demonstrate the step-by-step parsing process using a Shift-Reduce parser for the given input string. Provide the sequence of shifts and reductions, including the state of the parsing stack at each step. Make clear how the parser proceeds from one step to the next and whether it shifts or reduces, following the grammar rules. **OR** Explain the role of parallel corpora in training machine translation models. Also, highlight it's contribution in the translation quality of the system along with the use of "post-editing".		