Instructions to the students

1. **The students have to write the following in the answer sheet**

SASTRA DEEMED TO BE UNIVERSITY
SCHOOL OF COMPUTING
END SEMESTER EXAMINATION - May 2025

B. Tech - VI-SEMESTER

Name:

Register number:

Subject code: CSE405R02

Subject name: NATURAL LANGUAGE PROCESSING

Date: 07-05-2025

Time: 11:30 AM-2:30 PM

2. **Two questions are provided, each question carries 25 marks**

S.No.	Q. No.	Description	Max Marks	Marks Given
1	Question 1	Aim & Procedure	10	
2		Program	10	
3		Output	5	
4	Question 2	Aim & Procedure	10	
5		Program	10	
6		Output	5	
7		Total	50	

Examiner 1 **Examiner 2**

3. **Students have to use the dataset attached**
4. **Students have to verify the required packages within the first 15 minutes**
5. **Finally, upload the answers without fail**
6. **No additional sheets will be provided**

SASTRA Deemed University

Department of Computer Science and Engineering

Natural Language Processing (CSE405R02)

End Semester Semi Theory Semi Lab Examination

07.07.2025 **11:30AM-2:30PM**

1.

a. Read the WSD_light.xlsx file. Tokenize the corpus and clean the data (remove stop words, punctuation, etc.)

b. Compute prior probabilities for each sense category and conditional probabilities for each feature given the sense using Laplace smoothing.

c. Apply the Bayesian Disambiguation (WSD) model to classify the ambiguous word "light" in various contexts and output the predicted sense for the test sentences containing "light".

2.

Design a feed-forward neural network to perform text classification. Use Bag of Words and TF-IDF vectorization techniques. Dataset: moviereview.txt.

Show the

i. Display the label encoding of the document.

ii. Accuracy of BOW and TF-IDF

iii. Display the Classification report

iv. Display the confusion Matrix