



QP CODE: 25804509



25804509

Reg No :

Name :

INTEGRATED M.Sc DEGREE EXAMINATION, OCTOBER 2025

Ninth Semester

INTEGRATED MSC COMPUTER SCIENCE-DATA SCIENCE

CORE - ICSD9CR1 - TEXT ANALYTICS & NATURAL LANGUAGE PROCESSING

2021 ADMISSION ONWARDS

22EFECF1

Time: 3 Hours

Weightage: 30

Part A (Short Answer Questions)

*Answer any **eight** questions.*

Weight 1 each.

1. Differentiate NLU and NLG.
2. Define Named Entity Recognition.
3. Give an example of how lists can be used to store tokens in an NLP program.
4. Demonstrate add and access elements in a dictionary.
5. Explain the corpus disk structure and describe best practices for managing large text corpora.
6. List tools used for reading processed corpus files.
7. Define cross-validation.
8. Discuss the importance of model deployment in real-world NLP.
9. Illustrate the use of scatter plots to visualize feature spaces in NLP tasks.
10. Compare hierarchical clustering and K-means clustering in terms of visualization clarity.

(8×1=8 weightage)

Part B (Short Essay/Problems)

*Answer any **six** questions.*

Weight 2 each.

11. Explain word sense disambiguation with example.
12. Explain Zipf's Law.
13. State the use of CountVectorizer() function.
14. List key points for checking data quality in a corpus.
15. Define multi-label classification with a simple example.





16. Explain what model retraining means in production.
17. Describe one weakness of n-gram models in text generation.
18. Describe how to show a classifier's decision boundary visually.

(6×2=12 weightage)

Part C (Essay Type Questions)

*Answer any **two** questions.*

Weight 5 each.

19. Summarize the integration of regular expressions with NLTK pipelines for efficient preprocessing.
20. Explain why intermediate storage of preprocessed corpus data is useful.
21. Discuss how cross-validation improves the reliability of model evaluation in text classification.
22. Explain how n-gram range affects feature space and suggest a simple method to select n-grams.

(2×5=10 weightage)

