

|  |  |  |
| --- | --- | --- |
| **Trimester: March 2025**  **Examination: End term Examination** | | |
| **Program code: 01**  **Program: MBA BADS Major** | **Class: SY** | **Trimester: VI**  **(SVU 2024)** |
| **Name of the Constituent College:**  **K. J. Somaiya Institute of Management** | **Name of the department/Section/Center: Dept of Data Science and Technology** | |
| **Course Code: 217P01C615** | **Name of the Course: NATURAL LANGUAGE PROCESSING** | |

**Maximum Marks: 50 Date: 08/04/2025**

**Duration: 3HRS**

**Instructions: - Q1 is compulsory, attempt any four out of remaining**

|  |  |  |
| --- | --- | --- |
| **Question No.** |  | **Max.**  **Marks** |
| Q 1 | Web scrape any web page with atleast 5 column data from atleast 3 pages in a pagination scenario.  Web scrape any Wikipedia html table  Save both the datasets in csv file. | 10 |
| Q2 | How would you describe the steps involved in an NLP pipeline, provide case-based examples from real-world business scenarios to demonstrate how each step contributes to solving business problems or enhancing decision-making? | 10 |
| Q3 | How would you explain the various phases of NLP pre-processing as a key process in an NLP project within a business organization, and what is its significance for enhancing business operations and decision-making? | 10 |
| Q4 | Explain how TextBlob performs sentiment analysis, focusing on polarity and subjectivity. Provide a business example where TextBlob could be applied to analyze customer reviews. Additionally, compare VADER with TextBlob, highlighting how VADER handles sentiment analysis differently, especially for short texts or social media. Use an example of a tweet or short sentence to demonstrate VADER’s effectiveness. Discuss the advantages of using VADER in a business context, such as monitoring brand sentiment on social media. | 10 |
| Q5 | Difference between any two:   1. Syntactic vs. semantic analysis (one example each) 2. Data cleaning vs. Data preprocessing 3. NLG vs NLU | 10 |
| Q6 | Explain NER and POS Tagging, bigram trigam and ngram for an regional language business scenario, elaborate its applications in business or research domain. | 10 |
| Q7 | How would you approach feature engineering in an NLP project, and what techniques would you apply to extract meaningful features from text data to improve model performance in a business context? Implement minimum 3 lexical feature engineering techniques. | 10 |
| Q8 | Preform sentiment analysis and emotional analysis on movie\_review dataset of nltk package, give minimum 5 key business insights from the same. | 10 |