**NLP - Topic Modeling**

**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_ Batch ID:** \_\_\_\_\_\_\_\_\_\_\_

**Topic: NLP - Topic Modeling**

**Grading Guidelines:**

**1. An assignment submission is considered complete only when correct and executable code(s) are submitted along with the documentation explaining the method and results. Failing to submit either of those will be considered an invalid submission and will not be considered for evaluation.**

**2. Assignments submitted after the deadline will affect your grades.**

**Grading:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ans** | **Date** |  |  | **Ans** | **Date** |
| Correct | On time | A | 100 |  |  |
| 80% & above | On time | B | 85 | Correct | Late |
| 50% & above | On time | C | 75 | 80% & above | Late |
| 50% & below | On time | D | 65 | 50% & above | Late |
|  |  | E | 55 | 50% & below |  |
| Copied/No Submission |  | F | 45 |  |  |

* **Grade A: (>= 90):** When all assignments are submitted on or before the given deadline.
* **Grade B: (>= 80 and < 90):** 
  + When assignments are submitted on time but less than 80% of problems are completed.

(OR)

* + All assignments are submitted after the deadline.
* **Grade C: (>= 70 and < 80):** 
  + When assignments are submitted on time but less than 50% of the problems are completed.

(OR)

* + Less than 80% of problems in the assignments are submitted after the deadline.
* **Grade D: (>= 60 and < 70):**
  + Assignments submitted after the deadline and with 50% or less problems.
* **Grade E: (>= 50 and < 60):** 
  + Less than 30% of problems in the assignments are submitted after the deadline.

(OR)

* + Less than 30% of problems in the assignments are submitted before the deadline.
* **Grade F: (< 50):** No submission (or) malpractice.

**Hints:**

**1. Business Problem**

* 1. **What is the business objective?**
  2. **Are there any constraints?**

**2. Data Pre-processing**

**2.1 Data Cleaning, Feature Engineering, etc.**

**3. Exploratory Data Analysis (EDA).**

**4. Model Building**

**4.1 Perform Data Cleaning, Stemming, Lemmatization, Topic Modelling and Text Summarization.**

**5. Write about the benefits/impact of the solution - in what way does the business (client) benefit from the solution provided?**

**Problem Statement-1**

1. Perform NLP – Topic Modelling and Text Summarization by following all the steps as mentioned below.
2. Data Cleaning using regular expressions, Countvectorizer, POS Tagging, NER, Topic Modelling (LDA, LSA) and Text Summarization.

Hint: - Use Data.csv file given in hands-on material.

Text

Description automatically generated

**Problem Statement-2**

Perform Topic Modelling and Text Summarization on the given text data. Use the NLP-TM text file.

**A picture containing letter

Description automatically generated**