# Topic: Text Mining (NLP)



**Instructions**

Please share your answers filled inline in the word document. Submit Python code and R code files wherever applicable.

Please ensure you update all the details:

**Name: Batch Id:**  **Topic: Text Mining and NLP.**

1. **Business Problem**
   1. **Objective**
   2. **Constraints (if any)**
2. **Work on each feature of the dataset to create a data dictionary as displayed in the below image:**



**2.1 Make a table as shown above and provide information about the features such as its Data type and its relevance to the model building, if not relevant provide reasons and provide description of the feature.**

**Using R and Python codes perform:**

1. **Data Pre-processing**

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

**3.2 Outlier Imputation**

1. **Exploratory Data Analysis (EDA):**
   1. **Summary**
   2. **Univariate analysis**
   3. **Bivariate analysis**
2. **Model Building**
   1. **Extract textual data from Websites such as Amazon, Snapdeal, IMDB, Twitter**
   2. **Clean text data and build Word Cloud for both Positive and Negative words as well as perform Sentimental Analysis**
   3. **Briefly explain the model output in the documentation**



1. A screenshot of a cell phone

   Description automatically generated**Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided**

# Note:

**The assignment should be submitted in the following format:**

* **R code**
* **Python code**
* **Code Modularization should be maintained**
* **Documentation of the model building (elaborating on steps mentioned above)**

**Standard Grading Guideline :**

Grade A: All assignments submitted correctly on Time (with all mentioned content like: python code, r code, documentation)

Grade B: All assignments submitted but post the due date. Or Partial assignments are submitted.

Grade C and Grade D: Partial assignments submitted with incorrect answers, or worked on only R or Python or not all the content is submitted.

Grade F: Partial assignments submitted with incorrect answers and not all the content is submitted.

**Problem Statement: -**

In the era of widespread internet use Business now a days require to be ahead of their competitors to this they would like to analyze the performance of their product in the market which in turn will help them to improve their marketing and product in the future. Help the business grow by analyzing the reviews of their product by using Text mining and sentiment analysis.

Task 1:

1. Extract reviews of any product from ecommerce website like Snapdeal and Amazon
2. Perform sentimental analysis prepare Word Cloud for Unigram and Bigram.

Task 2:

1. Extract movie reviews for any movie from IMDB and perform sentimental analysis
2. Extract anything you choose from the internet and do some research on how we extract using R and Python Programming and perform sentimental analysis.