# Topic: Text Mining (NLP)

**Hints:**

1. **Business Problem**
   1. **What is the business objective?**
   2. **Are there any constraints?**
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. And if not relevant, provide reasons and a description of the feature.**

1. **Data Pre-processing**

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

**3.2 Outlier Treatment**

1. **Exploratory Data Analysis (EDA):**
   1. **Summary**
   2. **Univariate analysis**
   3. **Bivariate analysis**
2. **Model Building**
   1. **Extract text data from websites such as Amazon, Snapdeal, IMDB, Twitter, etc.**
   2. **Clean the data and build a word cloud for both positive and negative words. Perform Sentiment Analysis as well.**
   3. **Briefly explain the model output in the documentation.**
3. **Write about the benefits/impact of the solution - in what way does the business (client) benefit from the solution provided?**

**Problem Statement: -**

In the era of widespread internet use, it is necessary for businesses to understand what the consumers think of their products. If they can understand what the consumers like or dislike about their products, they can improve them and thereby increase their profits by keeping their customers happy. For this reason, they analyze the reviews of their products on websites such as Amazon or Snapdeal by using text mining and sentiment analysis techniques.

Task 1:

1. Extract reviews of any product from e-commerce website Amazon.
2. Perform sentiment analysis on this extracted data and build a unigram and bigram word cloud.

Task 2:

1. Extract reviews for any movie from IMDB and perform sentiment analysis.

Task 3:

1. Choose any other website on the internet and do some research on how to extract text and perform sentiment analysis