**Data Analytics Laboratory**

**Task 6**

**Text Analytics: Sentiment Analysis**

**Introduction**

* Text analysis, sometimes called text analytics, refers to the representation, processing, and modeling of textual data to derive useful insights.
* The key to effective text analytics lies in employing appropriate data preprocessing techniques.
* Collect Raw text: Social network application programming interfaces (APIs) process data feeds, or scrape pages and use product names as keywords to get the raw data.
* Represent text: Convert each review into a suitable document representation with proper indices, and build a corpus based on these indexed reviews.
* The usefulness of each word is estimated using TFIDF.
* Identify whether the reviews are positive or negative. Review the results and gain greater insights

Diagram

Description automatically generated

**Prerequisites**

1. What are stop words? Why we should not consider stop words while performing text based analytics?

2. What is the difference between stemming and lemmatization?

3. Why do we use inverse document frequency? What is its output represents?

4. List some of the application where this sentiment analysis can be applied?

**Exercise**

**Perform sentiment analysis for the attached IMDB movie dataset and predict each into two category (Positive or Negative review)**

**Link to Download the dataset:** <https://bit.ly/2X2ET43>

**Results**

The program is implemented in python and the output is observed.

**Faculty Signature**