**NLP- Topic Modelling**

**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: \_Shubham Sharma \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Batch Id: \_\_\_:**  DS WDMCO 050121 Blr **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Topic: NLP- Topic Modelling**

**1. Business Problem**

* 1. **Objective**
  2. **Constraints (if any)**

**2. Python codes perform:**

**3. Data Pre-processing**

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

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

**5. Model Building**

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

**6. 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:

* Python code
* Documentation of the modules (elaborating on steps mentioned above)

**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, Count Vectorizer, POS Tagging, NER, Topic Modelling (LDA, LSA) and Text summarization.

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

***INFERANCES :-***

* **For solving the problem statement I used the following steps.**
* **At first the set extraction has performed with only holding the text from the data.**
* **Then data cleansing is been performed using regular expression library and “nltk” library.**
* **Then for LDA model “GENSIM” library is used and the plot for LDA model has been done by using matplotlib.**
* **Then coming to LSA model the “ GENSIM “library is used and matplotlib for the visualization.**
* **While continuing on LSA the libraries like Beautifulsoup and pathlib is used.**
* **Finally coming to text summarizing the libraries like collection , string ,heapq are used and the program has been done successfully.**

Text

Description automatically generated

Problem Statement-2

Perform topic modelling and text summarization on the given text data hint use NLP-TM text file.

**A picture containing letter

Description automatically generated**

***INFERANCES :-***

* **For solving the problem statement I used the following steps.**
* **At first the set extraction has performed with only holding the text from the data.**
* **Then data cleansing is been performed using regular expression library and “nltk” library.**
* **Then for LDA model “GENSIM” library is used and the plot for LDA model has been done by using matplotlib.**
* **Then coming to LSA model the “ GENSIM “library is used and matplotlib for the visualization.**
* **While continuing on LSA the libraries like Beautifulsoup and pathlib is used.**
* **Finally coming to text summarizing the libraries like collection , string ,heapq are used and the program has been done successfully.**