**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: Ayyappa Bharthwaj Nukala**

**Batch Id: DSWDMCOS 21012022**

**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.

Text

Description automatically generated

Sol:

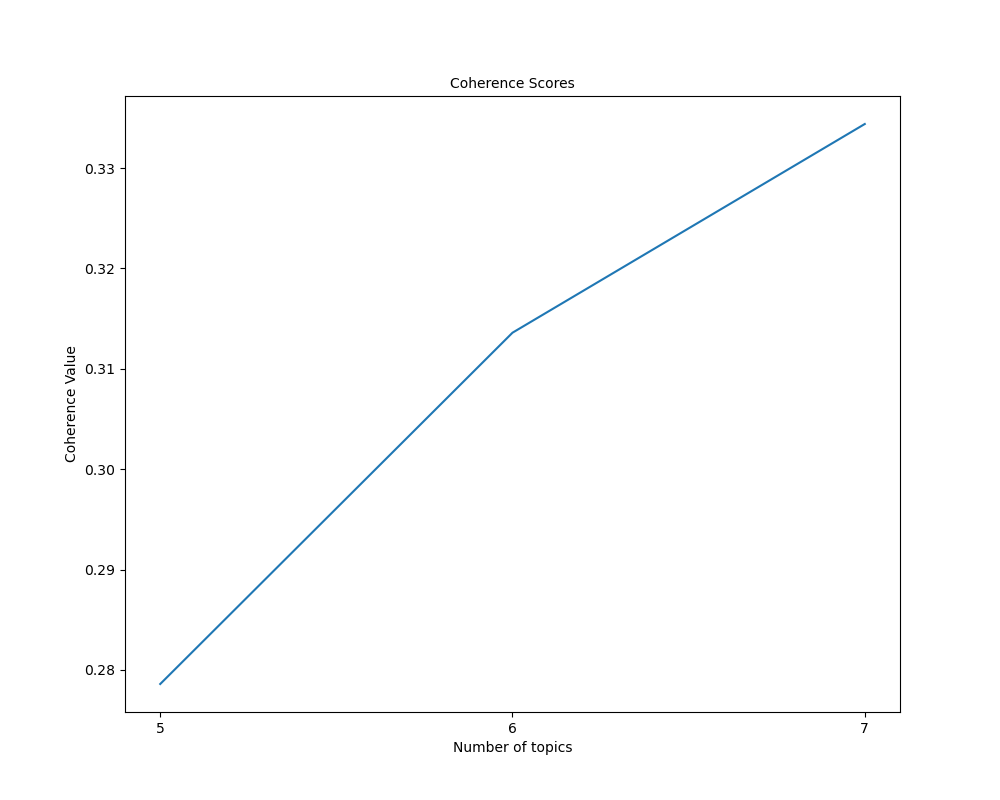
**Business Objective:** To create a topic modelling and text summarization for the tweets data

**Data Type:** The given are the tweets given by the different people having the particular tweet I’d

**Data Pre Processing:** I have done the data pre-processing on the data set in Python using custom functions. I have removed the punctuation marks, special character’s, Tokenization etc. After cleaning the data I have done the analysis on the data.

**Topic Modelling:** I have done the LDA (Latent Direchlet Allocation) for the tweets data and I have made the model for the same using Python.

**Text Summarization:** Again I have cleaned the complete data set and then I have done the tokenization, removing the stop words, bag of words. After cleaning the data I have created a text summarization model and by using that I have summarized the data in Python.



Problem Statement-2

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

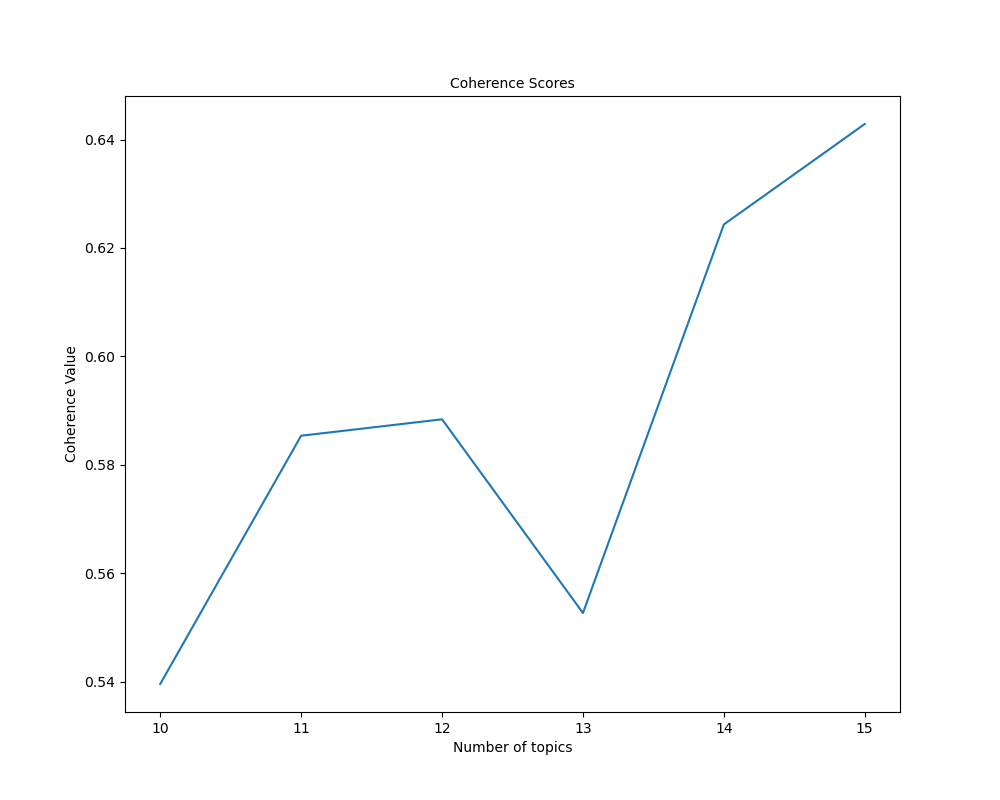
**Business Objective:** To create a topic modelling and text summarization for the text data.

**Data Type:** The text data of an article is given.

**Data Pre Processing:** I have done the data pre-processing on the data set in Python using custom functions. I have removed the punctuation marks, special character’s, Tokenization etc. After cleaning the data I have done the analysis on the data.

**Topic Modelling:** I have done the LDA (Latent Direchlet Allocation) for the text data and I have made the model for the same using Python.

**Text Summarization:** Again I have cleaned the complete data set and then I have done the tokenization, removing the stop words, created bag of words. After cleaning the data I have created a text summarization model and by using that I have summarized the data in Python.



**A picture containing letter

Description automatically generated**