# **Milestone report Capstone Project**

**Project Title**:

**Text Summarization.**

Project Proposal:

What is the problem to be solved?

**Text Summarizer is a NLP based problem which is used to convert huge text into small Summary**.

**Who is this project beneficial for?**

**Our client are student, Scientist, News media like inShort.**

## **What data are you using? How will you acquire the data?**

**In this the user has to provide a Wikipedia Url which will be summarized for the user.**

## **How you’ll solve this problem?**

1. **Data Wrangling**: converting the text in proper structure like removing the unwanted words, lemmatization, removing stop words.

**2.** **Exploratory Data Analysis:** i will plot relation between different parameters in the dataset and based on that and based on that I can find the trends for storytelling.

**3.** **Inferential Statistics:**  build statistical relationship in the dataset.

**4.** **Machine Learning Model:** using Page Rank Technique to find the most important statements from the text.

## **What are your deliverables?**

I will use the following tool to work on my project.

1. Jupyter Notebook

2. Python 3.7

3. Google doc for creating the document

4. Powerpoint for data Storytelling presentation

# **Data Wrangling Report:**

### **1.** **Extract Text:**

The first step which I did in this project was to collect the p tag from the url given by the user using Beautiful soup.

### **2.** **Clean Text.**

The second step was 1. Removing \n from the string 2. Removing square brackets 3. Remove everything except brackets 4. Remove extra spaces.

### **3.** **Text Processing:**

The next step is to 1.tokenize sentences 2. Converting into lower case 3. Removing stopwords.

**4.** **Convert Into vectors:**

Convert into vectors.

### **5.** **Similarity matrix**

Create similarity matrix for each sentence.

Now the data is mostly clean and ready for the further process of Exploratory Data Analysis, Data Visualization, Machine Learning Model.

# **Exploratory Data Analysis And Storytelling:**

### **v ariables that are significant in explaining the project question?**

How the Recommendation will affect different variables into consideration?

There are many variables/ features which are taken into consideration for that is responsible for the Recommendation of a Movie.

These are as follows:

1. Popularity: One of the ways of recommending movie to the user is by popularity. If the movie is popular then we should recommend the movie to the user.

Ex. Titanic is the most popular movie of all time then we should definitely recommend this movie to the user.

2. Vote Count: More the vote that means that majority of the people have liked the movie so in that case if a movie has some good amount of vote count then we should recommend that movie to the user.

3. Vote Average: Similar to vote count even vote average plays an important role. Good vote average value means good movie so we should recommend that movie to the user.

4. Cast, keyword, genres, director: if the user has seen a movie then depending upon the cast, director, genres of the movie we can recommend the movie to the user which are most similar to the movie user has seen.

### **C orrelation between a pair of variables.**

What does the pair of variables suggest when compared with each other?

1. Budget and revenue: Both are positively correlated with each other if one increases the other one will also increase.

2. Popularity and vote count: Both are positively correlated with each other if one increases the other one will also increase.

3. The vote count and revenue: Both are positively correlated with each other if one increases the other one will also increase.