**Trends from News Headlines Project Proposal**

**GitHub Repository**:- <https://github.com/dhathrigundum/Group-11-NLP_news-headlines>

**Teammates**:

Sai Dev Prakash Janapareddi, ID: 11507630

Dhathri Gundum, ID: 11442139

Sri Harsha Swaraj Nadendla, ID:11501903

**Goals and Objectives: -**

* **Motivation: -**Lately, digitalisation has been the primary goal of the world, because of which the amount of data in every domain, be it State of art neural networks, Google’s search engine data, News data, Cancer data, Tweets data, every possible domain one can imagine has almost un-maintainable amount of data. One such domain where the data over past decades or few couple of decades ago is humongous in size and which provides the possibility to classify and understand what has happened in the world during a period of time is News data. This has been our main motivation. To try understanding and bringing out observations about the events that took in the world a decade ago. Well the time period is nothing specific, but we chose a decade as it provides us with enough data to classify and also not so huge to make the task tedious. This further opened the path to analyse the sentiment of the data, rather than just observations. When we dug deep, many of the headlines in the past, in the name of “catchy” headlines, are almost context less and ambiguous. This limitation has indirectly led meaningful news to become mere stop words in some cases. So, this limitation we observed has driven us to also try obtaining meaningful and sense making headlines when not-so-meaningful headlines are given as input. Overall, thanks to Natural Language Processing, we are i)classifying the text in News dataset obtained over a decade, ii)analyse the sentiment in it, and iii)bring out meaningful or rather sense-making headlines(text basically).
* **Significance: -** Classification of text(news) and placing in a category makes a lot of sense to understand the data and observe trends, which further leads to many questions and new ideas. Understanding this trend can be useful in many areas like for example: in stocks analysing the previous information about a stock its position in each quarter of a year and predicting the market and its effect from other competitors. The major significance what we believe is, ambiguous headlines will now make sense. When this kind of research is done or projects are made, all the domains which run on the statistics in the past, based on observations and these trends benefit a lot.
* **Objective: -** To clean the data obtained and pre-process it to make it easy to use in the code. Understand the news, history or such major events occurred in the past. Later classifying it as under which domain it comes, example: crime, event, war, etc. Then to bring out the shift in the sentiment of the news data, whether it has been positive or negative or neutral. We have chosen to code the entire project in python and utilize its libraries which have machine learning and NLP tools. Plotting necessary and insightful graphs for better visualisation of the outputs obtained.
* **Advancements:-** If time permits, we are very much excited about somehow, involving speech recognition or text to speech or vice-versa. If a random keywords about an event in history are given as input, the exact news headlines that match, the sentiment of the headline, and if the headline does not make complete sense, a meaningful headline that could have been an appropriate headline should be returned as an audio(text to speech).This is totally a speculated idea, if implementation of all the other features are done, we will try implementing this one too.
* **Features: -** The key features we would like to focus on are Analysing and classify the data related to news of over a decade, performing sentiment analysis on the data, and generating meaningful headlines. Along with those a few visualizations represent data that help categorize different domains of data.

**References: -**

[*https://www.kaggle.com/therohk/million-headlines*](https://www.kaggle.com/therohk/million-headlines)

[*https://towardsdatascience.com/how-i-used-natural-language-processing-to-extract-context-from-news-headlines-df2cf5181ca6*](https://towardsdatascience.com/how-i-used-natural-language-processing-to-extract-context-from-news-headlines-df2cf5181ca6)

[*https://neptune.ai/blog/exploratory-data-analysis-natural-language-processing-tools*](https://neptune.ai/blog/exploratory-data-analysis-natural-language-processing-tools)