

SPEECH DECODING & MAPPING BY THE NATURAL LANGUAGE PROCESSING LEADING TO AI

(A CASE STUDY OF HON'BLE PM SPEECHES)

RAMYA EMANDI
SANKARA RAO EMANDI

Ramya Emandi, works as a Programme Assistant at United Nations Development Programme, India (UNDP). She is the corresponding author for this paper. She is in a constant pursuit towards interdisciplinary/ pluralistic approach toward policy frameworks, trying to connect the dots from an engineer to manager to economist. She is an Engineer, MBA and an Advanced Studies graduate in Economics with an overall 8 years of experience in Corporate, Research and Multilateral Organisations. She is passionate about evidence-based policy & applied economics.

Sankara Rao Emandi, is an ex Managing Director and a CEO at public sector non-banking financial company IFCI Ltd. and Chairman at 5 other subsidiary organizations. He is an academician & researcher, a trained technocrat, a respected finance specialist and a prudent policy formulator. He has contributed immensely to the areas of banking in industry & infrastructure finance for over 30 years. He is an alumni of IIT Bombay (PhD), IIT Kharagpur (M. Tech), Pondicherry Central University (PGDBA), C.Engg and Andhra University (B.E. Electrical Engineering). The views expressed by the authors are independent and do not represent of any organizations.

Contact

ramya.emandi@undp.org; gkregk@gmail.com; Phone numbers: +919871593509; +919999443728

Speech Decoding & Mapping by The Natural Language Processing Leading to AI

(A Case study of Hon'ble PM speeches)

Artificial Intelligence (AI) has immense potential to improve the popularity in different spheres of human beings, businesses and empowering the global economy. Hon'ble Prime Minister had emphasized on the importance and the usage of Artificial Intelligence in his speeches. He mentioned that the young researchers and scientists will have to continuously work on a 'One Life One Mission' so that there is maximum use of drone technology, artificial intelligence technology and modern farm equipment in the country's agriculture. Amongst 350 start-ups that are being supported by the government, many of them are related to food processing, Artificial Intelligence, Internet of things (IOT), manufacturing of smart farming equipment and renewable energy. AI can be used effectively for the public policy designs and proper regulation in almost all sectors. The findings of McKinsey Global Institute have emphasized the fact that Artificial Intelligence will have immense impact on the global economy in the time to come. McKinsey identified five AI technologies that will gain a lot of attention in the time to come.

1. **Computer Vision**
2. **Natural Learning Processing (NLP)**
3. **Virtual Assistants**
4. **Robotic Process Automation**
5. **Advanced Machine Learning**

Artificial intelligence is basically the ability of a machine to replicate the intelligence or cognitive functionalities of a human being. AI enables machines or computer-enabled robotic systems to learn from experience, process information and generate outcomes in learning, decision making, and complex problem solving. Artificial intelligence encompasses multiple disciplines that are inspired from the knowledge of mathematics, statistics & computer science to build models, tools & software programs. These tools & programs are trained to help humans arrive at complex decisions by connecting through IOT devices and communicating through digital communication networks.

The most important advantage of using AI is the ability of it to use unstructured data – for instance, text, images and audio. In the recent times, the usage of quantitative data has increased enormously. With the development of machine learning, the computers can process non-quantitative data and enable humans to gain deeper insights. Some of the examples that showcase the competence and the extensive usage of AI are Fake News Detection, Disease Detection, Speech Emotion Recognition, Sentiment Analysis, Spam Classifier, Handwritten Digit Recognition, Chatbot, Driver Drowsiness Detection, Music Recommendations, Traffic Signs Recognition and Image Caption Generator.

In order to showcase one such example, we have worked on a ML-AI algorithmic speech decoding & mapping by the natural language processing.

A ML-AI Algorithmic Speech Decoding & Mapping By The Natural Language Processing

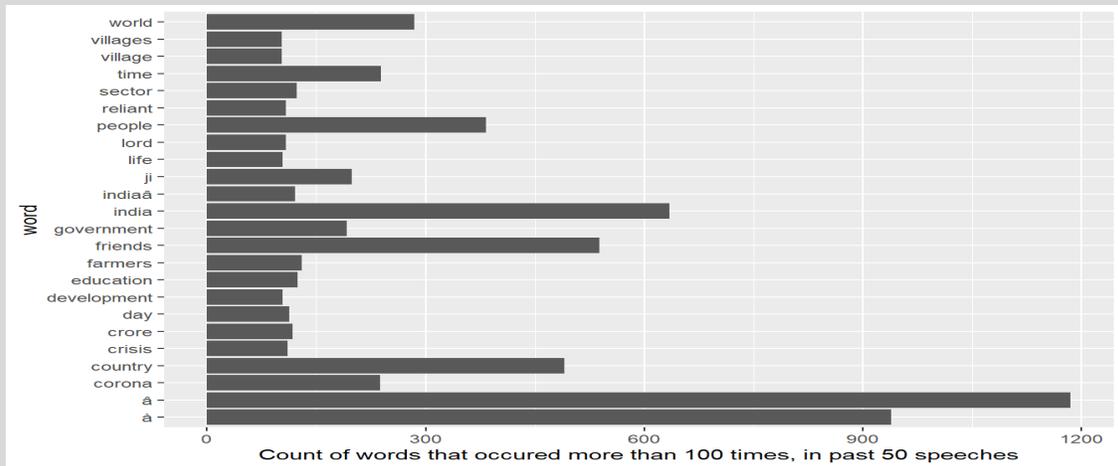
Objective: This is a study which highlights the importance of the Data Analytics and ML-AI simulations usage for the better understanding of the processes or activities by analyzing the Data in various forms such as Numbers,

Natural Language Text Processing, Voice or Pictures. For this study the Natural Language Processing has been considered.

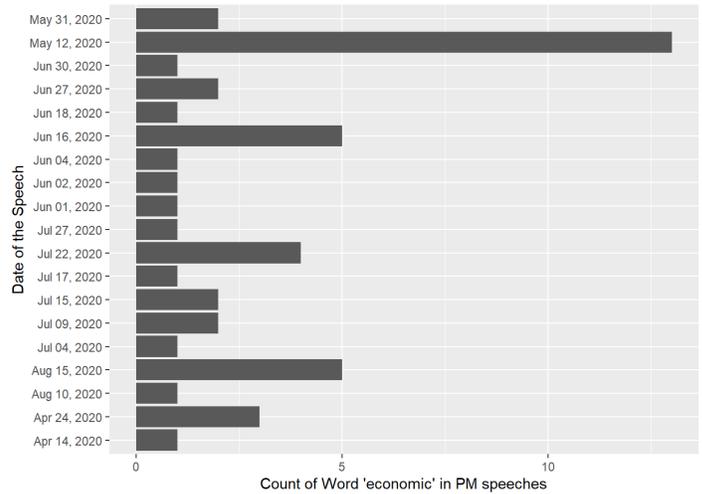
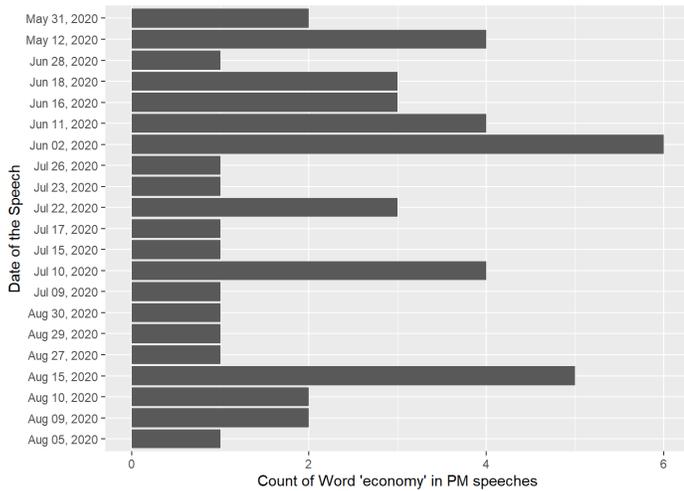
Case Study : For the above objective the Hon'ble PM last 50 speeches i.e. from Mar 2020 to Aug 2020 have been taken as the case study subject to look at how the importance of the word "Economy" has been used in the speeches in different contexts and changed over time by counting the number which times "Economy" mentioned in Hon'ble PM's speeches"

Data Analytics ML Tool : ML R-Programming Algorithm Design (RAG)

1. The past 50 speeches of PM are sourced from PMO archive has been considered using the web scrapping and converted them to the read-made CSV file format from Kaggle.com. It may be mentioned here that the past 50 speeches are from *14th April 2020 (1st) to Aug 30th 2020 (latest 50th)* Hon'ble PM speech.
2. As a part of the data compilation and structuring, firstly, the speech words were un-nested and each word is separated to be able to be analyzed. Secondly, all speeches were processed and analyzed by RAG counting the most used words. This processing gives a general usage of words used by Hon'ble PM in different contexts in his speeches in various functions. The below chart shows the words used more than 100 times respectively. It may be seen in the 50 speeches that the word "a" is used nearly 1200 times, similarly "India" has been used 650 times and word "friends" used 500 times



3. After the un-nesting the words, now the word ‘economy’ and ‘economic’ are considered as a sample word and analyzed by shortlisting the speeches out of 50 that have used these two words. Both these words could have been stemmed as one but, it may be noted that these words can be used distinctly and exclusively in completely



different contexts and hence stemming them into one is

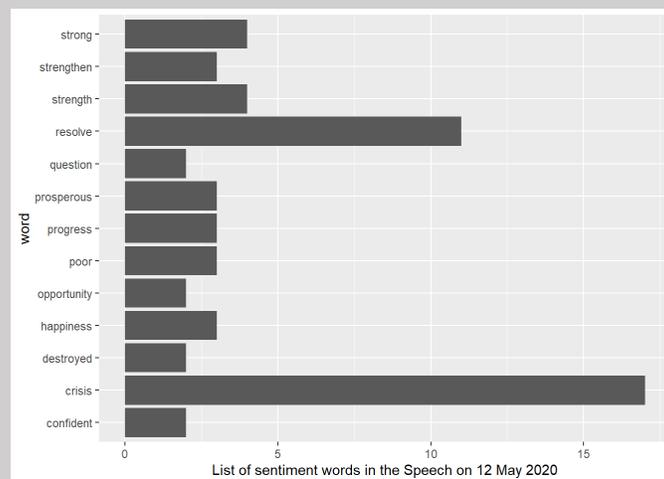
not favorable.

It is interesting to note that the word ‘economy’ wasn’t used before **May 12th** speeches. The focus on Economy started from **12th May, 2020** and was latest used in the speech on **5th Aug, 2020**. The speech on **12th May** is of relevance as the usage of word ‘economy’ and ‘economic’ has increased significantly.

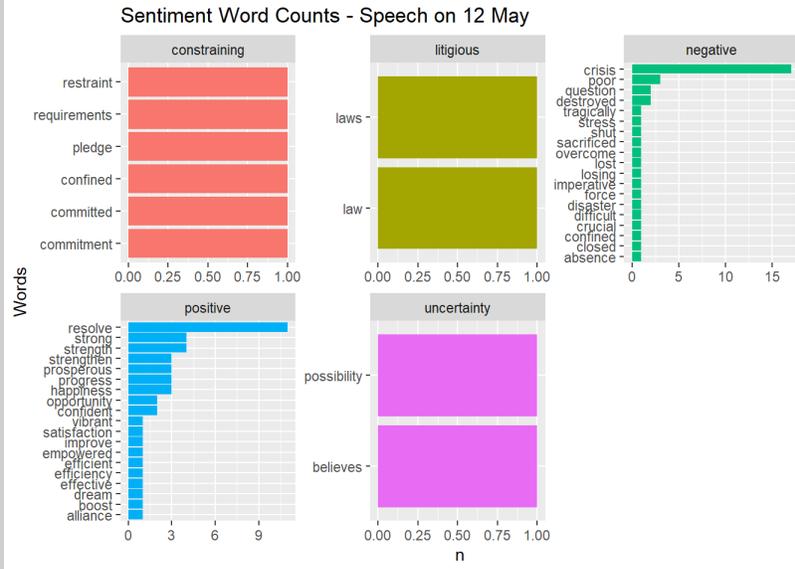
4. The speeches with maximum usage of word ‘economy’ are the speeches addressed on **2nd June 2020** and on **15th August, 2020 (Independence Day)**. Word count more than 4. In the case, word ‘economic’ is used at three occasions - 12 May 2020, 16 Jun 2020, 15 Aug 2020. The usage is highest (13) on **May 12, 2020**; should be interesting to analyze this speech as well.

A. 12th May 2020 speech dissection

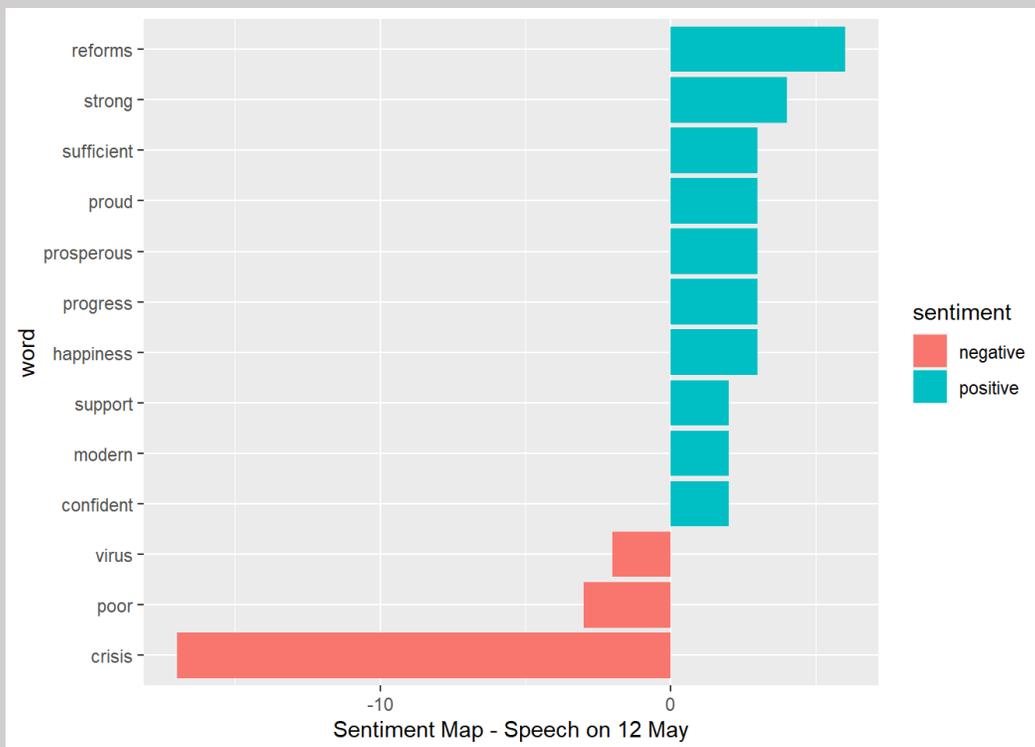
The words having certain level of sentiment in the speech addressed on 12th May, 2020 are listed below: The usage of word ‘crisis’ is evident, this was during the 3rd phase of lockdown.



Further these sentiment words are categorized into – positive sentiment, negative sentiment, contrasting sentiment, litigious sentiment and uncertain sentiment. The frequency and usage of positive sentiment is high but the usage of negative sentiment word ‘crisis’ is too high. The usage of uncertainty sentiment is minimal.



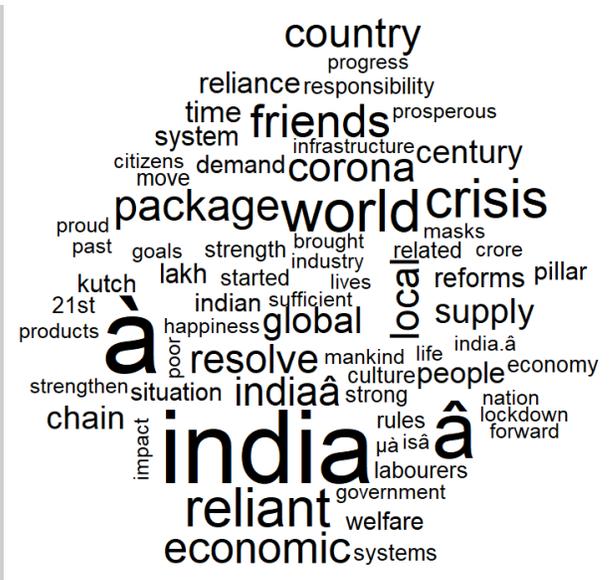
Overall sentiment is mapped – the speech has more positive sentiment words than the negative sentiment. However, the frequency of negative word is substantially high, especially the word ‘crisis’.



Visualisation: Word cloud for this speech, highlighting the frequent words as well as the sentiment words (colour cloud) is given below-

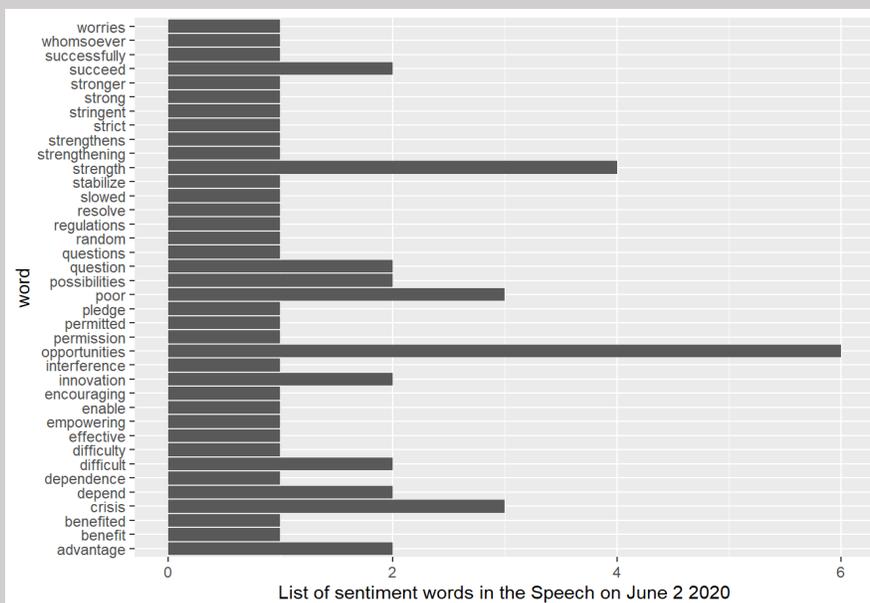
Among the negative words, the plight of the time is depicted – ‘virus’, ‘poor’, ‘crisis’

Among the positive words which are higher than the negative words, the response of government is depicted – ‘proud’, ‘reforms’, ‘happiness’, ‘sufficient’, ‘strong’, ‘progress’, ‘prosperous’. This is a hopeful and a positive speech.



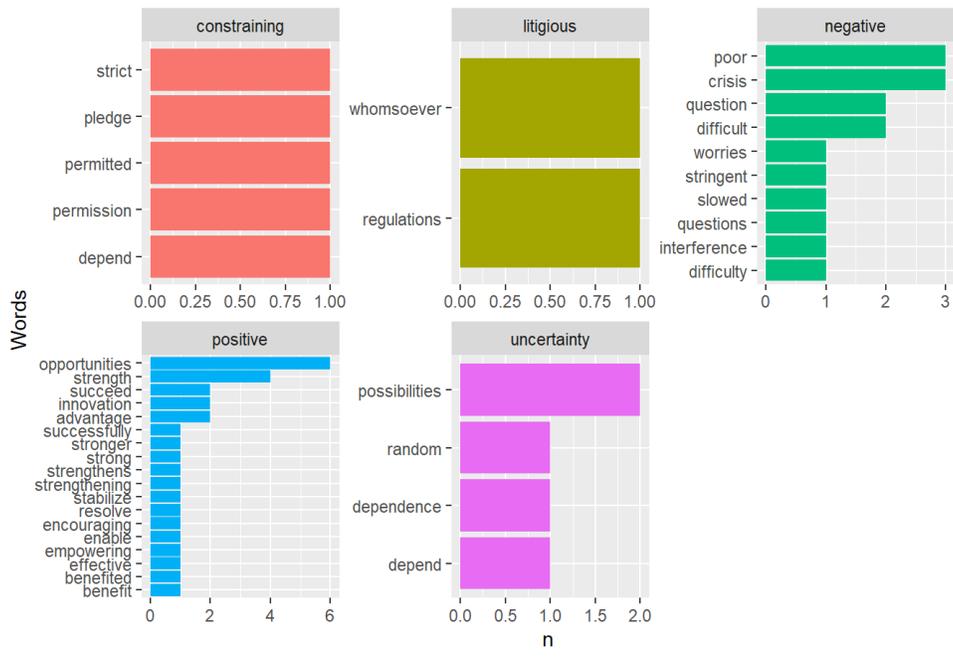
B. 2nd June 2020 speech dissection

The words having certain level of sentiment in the speech addressed on 2nd June are listed below:

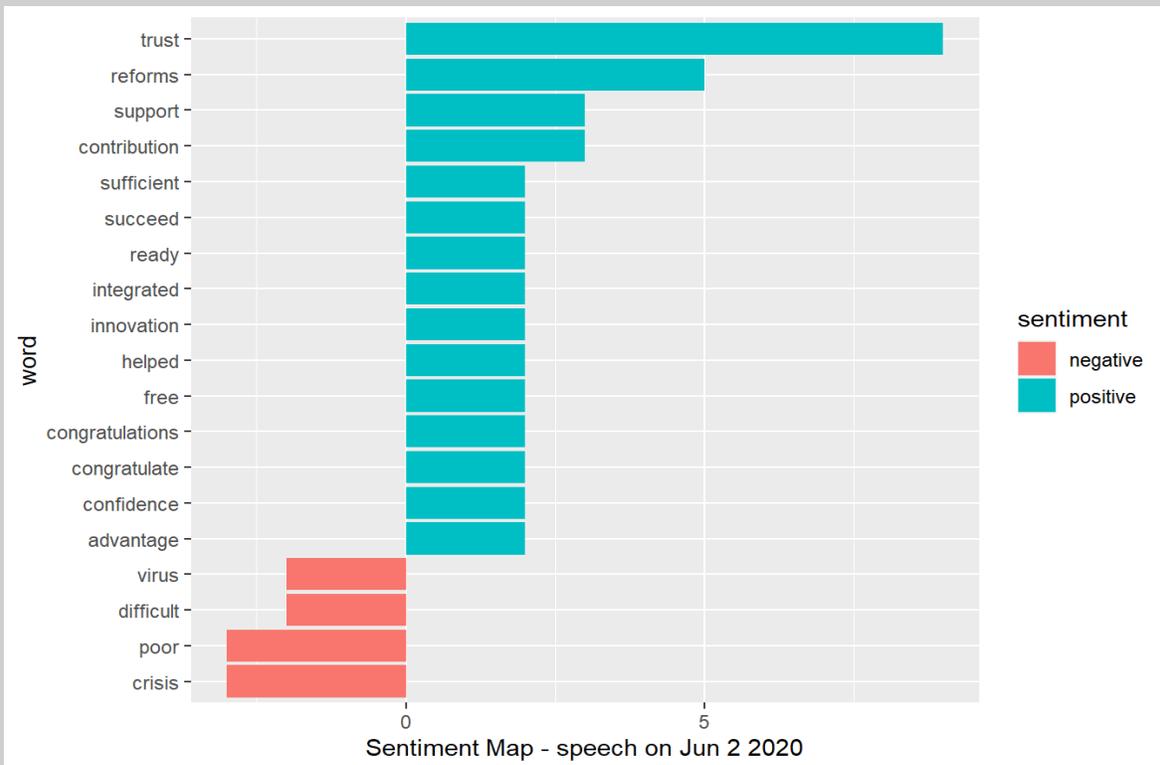


Further the above listed sentiment words are categorized into – Positive sentiment, Non-Positive sentiment, Contrasting sentiment, Litigious sentiment and Uncertain sentiment and segregated as show below respectively. It was observed that the frequency of usage of positive words is the most highest, as well, the number of positive sentiment words used in the speech are more than any other sentiment.

Sentiment Word Counts - Speech on Jun 2 2020



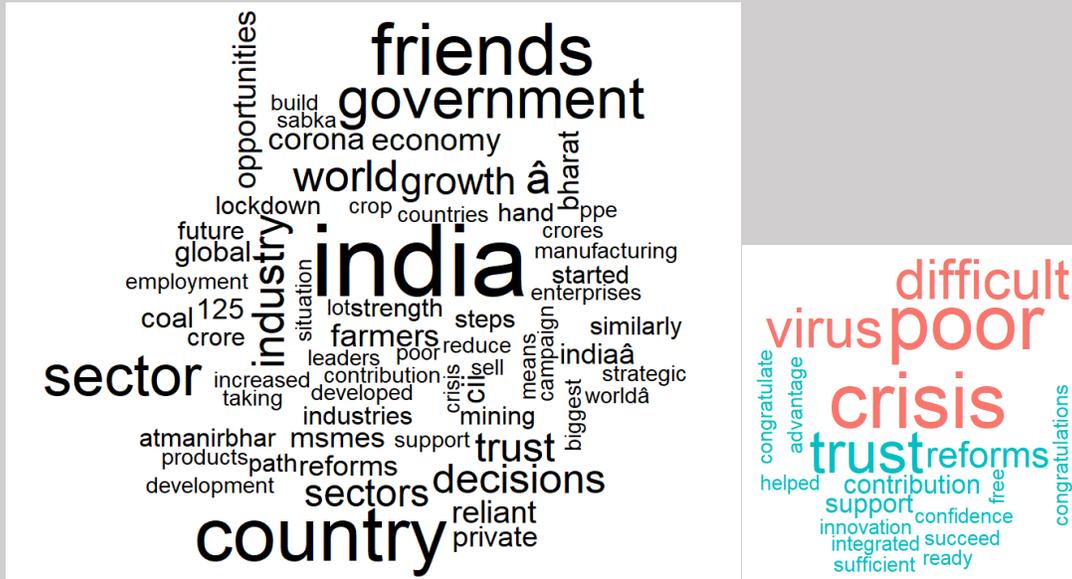
With the combination of all the mentioned sentiments had provided the Overall Sentiment Mapping (OSM). From the OSM map diagram below it can be seen and inferred that the speech is articulating more on a “positive note” and frequency of usage of positive words is also higher than the “non-positive” sentiment.



Visualization: The Word cloud for this 2nd June 2020 speech, highlighting the frequently used words as well as the sentiment words (colour cloud) is depicted below-

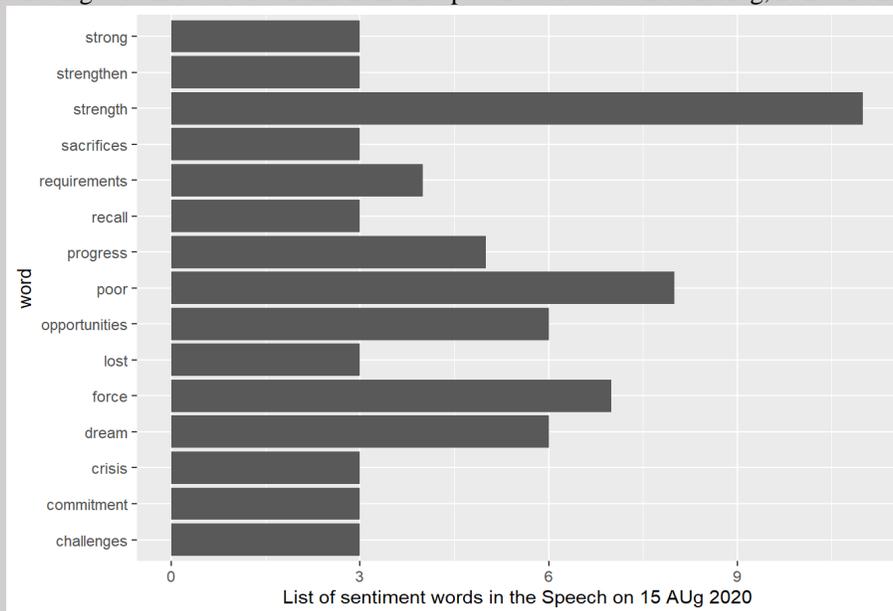
Among the **Non-positive** words, the plight of the time is depicted – ‘virus’, ‘poor’, ‘crisis’, ‘difficult’

Among the positive words, the response of government is depicted – ‘trust’, ‘reforms’, ‘contribution’, ‘support’



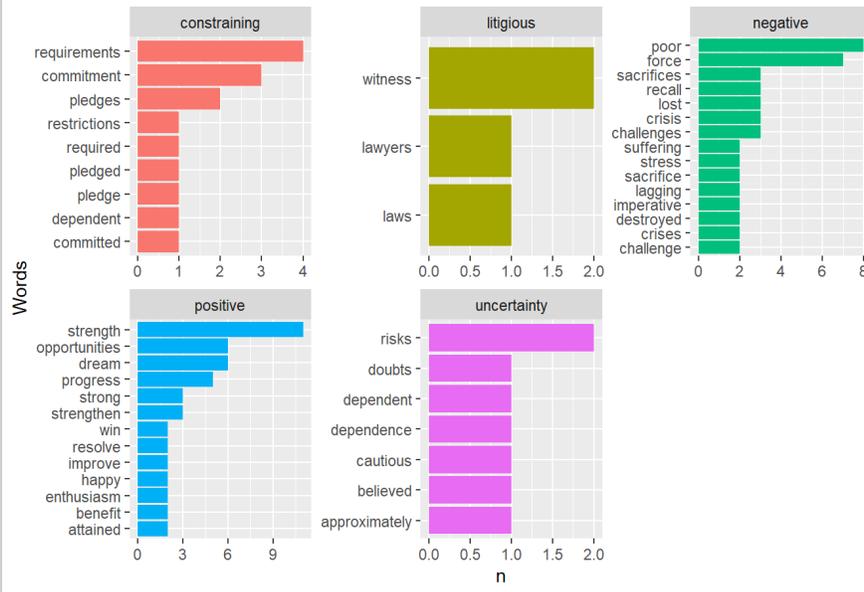
C. 15th Aug 2020 speech dissection

The words having certain level of sentiment in the speech addressed on 15th Aug, 2020 are listed below:

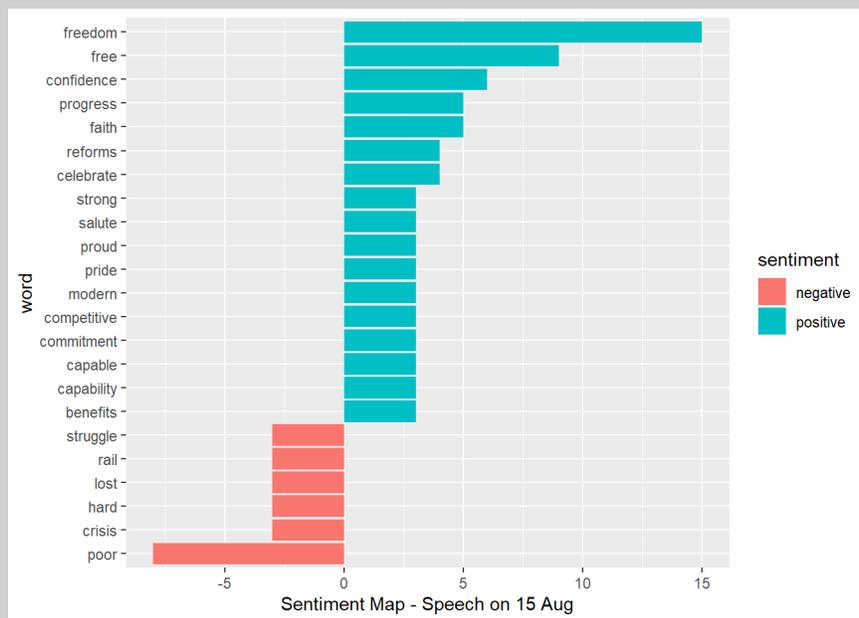


Further these sentiment words are categorized into – positive sentiment, negative sentiment, contrasting sentiment, litigious sentiment and uncertain sentiment. In this speech, the usage of both positive and negative sentiments is equally balancing. There seem to be equal and significant representation of constraining and uncertain sentiment.

Sentiment Word Counts - Speech on 15 Aug 2020



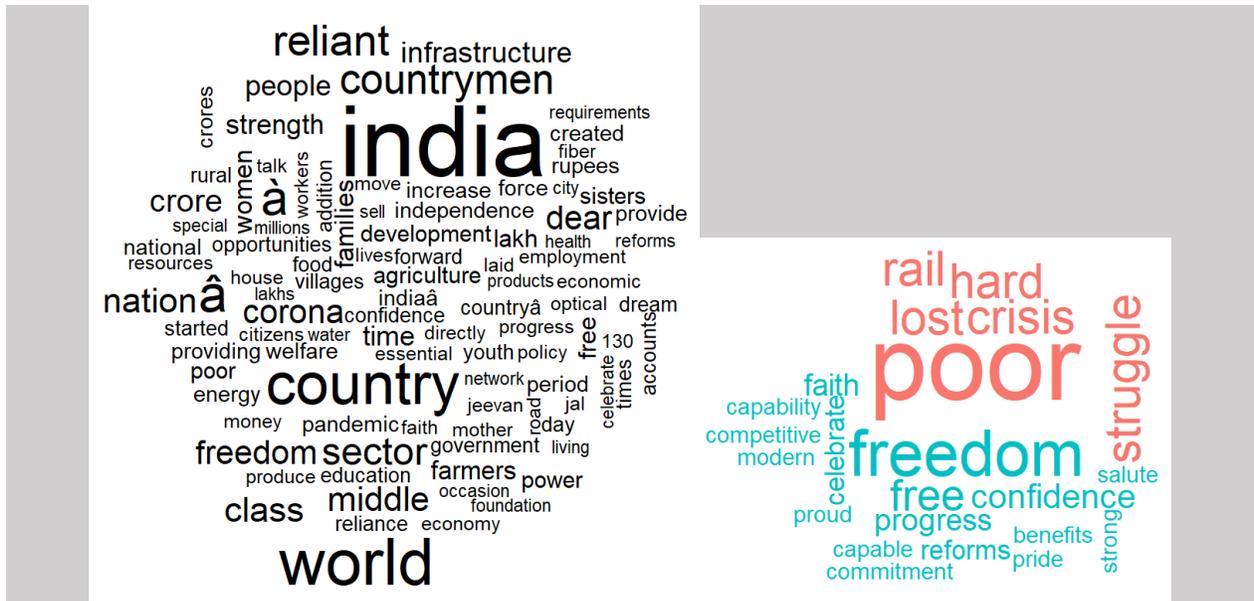
Overall sentiment is mapped – the speech is on a positive note and frequency of usage of positive words is also higher than the negative sentiment.



Visualisation: Word cloud for this speech, highlighting the frequent words as well as the sentiment words (colour cloud) is given below-

Among the negative words, the plight of the time is depicted – ‘lost’, ‘poor’, ‘crisis’, ‘hard’, ‘struggle’

Among the positive words, the response of government is depicted – ‘freedom’, ‘confidence’, ‘progress’, ‘celebrate’, ‘faith’



Conclusion – The World is changing, so is the data utility and processing. The decisions are more informed, and the perception of public is wide aware based on narrations. Earlier, dealing with text was not even considered in the applied statistical training of most disciplines. This is in direct contrast with how often it has to be dealt with prior to more common analysis, or how interesting it might be to have text be the focus of Data and AI analysis for better understanding of the trends and directions given by our policy leaders from time to time to have effective policy planning. This above classic example aims to provide a sense and visualization of the things one can do with text, and the sorts of analyses that might be useful.