

Understanding the importance of economy by analyzing Prime Minister Speeches

Abstract

Today India is ranked the fifth largest economy, and third largest in terms of Purchasing Power Parity (PPP). The Indian economy's GDP is pegged at \$ 2.9 trillion. With the adoption of the New Economy Policy in 1991, India saw a landmark shift in the economy with some top performing sectors such as:

- Agriculture
- Industry
- Service
- Food Processing
- Manufacturing

Beside the developments and boost into sectors, to get the best of the potential out proving to be helping hands in boosting the economy, a good governance between the central and state had helped coming up with initiatives such as:

- Goods and Services Tax (GST)
- Insolvency and Bankruptcy Code (IBC)
- Startup India
- Digital India

These among other have helped India climb 65 positions in the World Bank's Ease of Doing Business Index.

Source: Invest India Blogs

In this report, we are majorly focused on the speeches delivered by India's Prime Minister in the course of past 4 months, to understand the rise in importance of economy for the country. For the same, his media addressable transcript has been rendered, 50 in numbers working over unstructured data implementing machine learning using Python.

Keyword: Economy, PPP, sectors, initiatives, speeches, Python, machine learning

Background

With the idea of looking at how the importance of Economy has changed over time by counting the number of times "Economy" appeared in PM's speeches. I gazed over the past 50 speeches. The speeches are taken over from [PIB official site](#), from 6th March to 3rd September, 2020. List provided in the attachment. Since the focus is to understand the importance of economy, so the word economy does not only tell us in regard with the importance of India's economy, though many things combined in the essence of country's well being, carried forward from the past to ensure the country's progress in the coming years. The same thought process has been implemented for the thorough analysis of the PM's speech to extract the essence out of it. I have started off by taking the speeches into a pdf file format, an unstructured file, which in turn requires deep cleaning to understand its' whereabouts.

Data

Set of 50 pdf, all with the key phrase- speeches, as in the attachment was extracted from Press Information Bureau official website and stored in a folder for the analysis purpose.

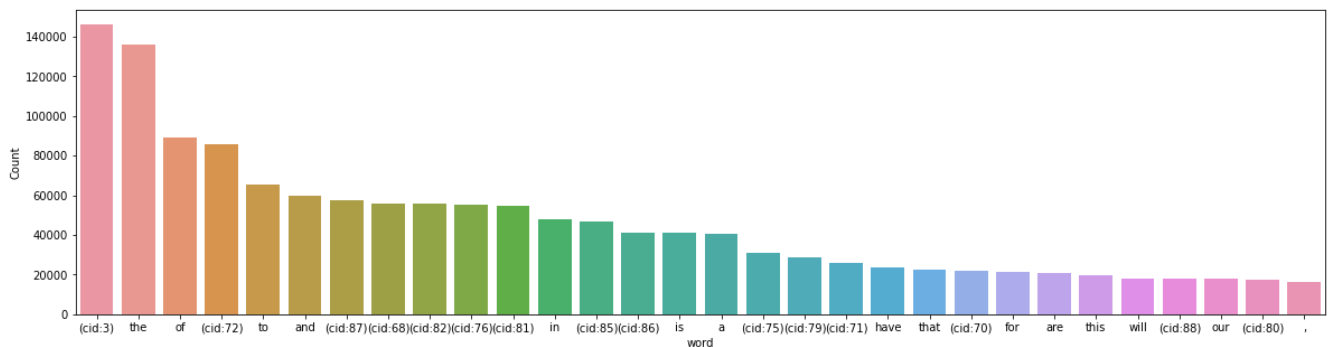
Analysis

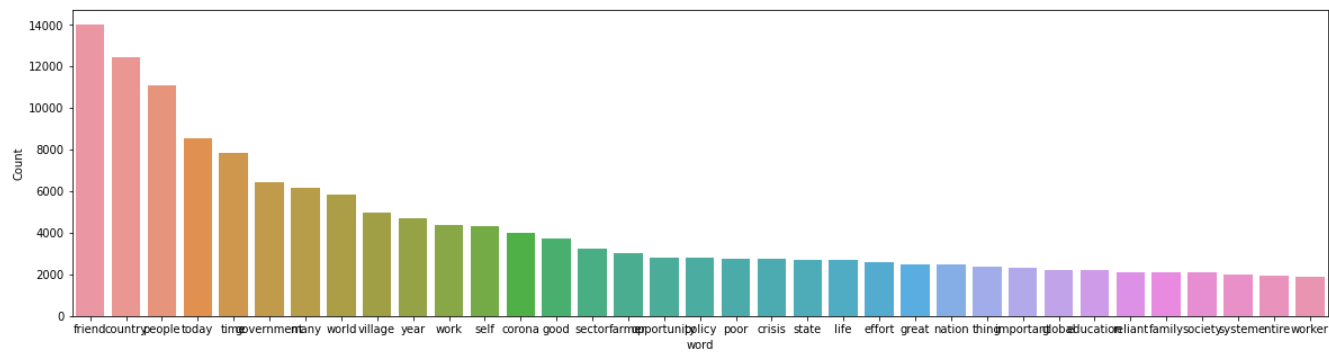
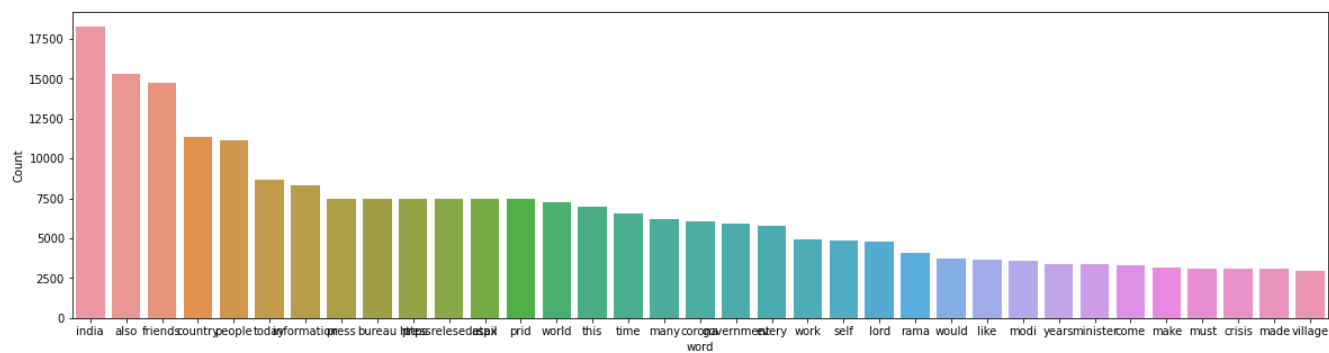
Methodology

Since the idea shifts from not only counting the word economy in the speeches to understand it's importance, rather than focusing on the essence of the whole speech converging to mediate our end goal. The project took place in 6 major steps with multiple sub steps within it. To name them;

- Packages and Data Import
 - Packages used are: **pdfminer3, pyLDAvis, nltk, spacy, matplotlib, pandas, numpy, gensim, seaborn**
- Data Pre processing
 - pdf data read into the python environment are rendered one by one and stored into a text file for ease of analysis.
 - Followed by importing the created text file for further analysis.
 - Our focus is to understand the topic related to the speeches in order to find association for the economy.
 - For the same we have used nltk, to clean the data, as spaces, redundant values are not analyzed by the machine.
 - We remove all the special characters, followed by removing the stop words such as is, are, the, kind of preposition or the most common occurring words
- Exploratory Data Analysis

see ispf
 people ∞
 US. Strange
 India
 diverse summit
 wonderful
 together





Topic 0

poor
year last youth
challenge
medical
business global
large world

Topic 1

important line
class
policy happy
government
work level
country
blessing

Topic 2

strong entire
economy
world
sector different
step year
thousand
infrastructure

Topic 3

major today
reform system
several
investment
decision
nation time
sector

Word Count and Importance of Topic Keywords

