Sentiment Analysis for News Covered in Print Media

Sentiment Analysis

Our project, "Sentiment Analysis for News Covered in Print Media," aims to analyze the sentiment of Indian political news articles. By leveraging NLTK, Natural Language Processing (NLP) techniques, we assess whether the content is positive, negative, or neutral. We also identify which political party is the focus of the article and determine the sentiment towards that party. The project involves extracting news articles using a News API key and utilizing the Beautiful Soup library for web scraping. By using Vader Sentiment, scores are predicted. This analysis helps in understanding media bias and public perception of political parties through print media.

Literature Review

• Antony Samuels, John Mcgonical 29 Nov, 2021

News Sentiment Analysis

https://drive.google.com/file/d/1-OQtyrlj5G2UTjVTWI_LvM4vClag3-7n/view?usp=drivesdk

The research paper focuses on the lexicon-based approach to sentiment analysis in news articles. It explores several key studies that have contributed to the field:

• **Lexicon-Based Sentiment Analysis:** The lexicon-based method relies on pre-defined dictionaries of words annotated with sentiment scores. Early studies, such as those by Turney (2002), emphasize the effectiveness of this approach in text classification, particularly in detecting sentiment polarity. Another key contribution is the work by Godbole et al. (2007), who implemented sentiment analysis on large-scale news and blog datasets using a lexicon-based approach, showcasing its scalability and effectiveness.

• **Challenges in Sentiment Analysis:** The literature identifies several challenges in sentiment analysis, including the ambiguity of words, context dependency, and the dynamic nature of language. These challenges often result in sentiment misclassification, which is a significant limitation of lexicon-based approaches.

• **Integration with Machine Learning:** While this paper primarily discusses lexicon-based methods, it acknowledges the growing trend of integrating these methods with machine learning algorithms to enhance accuracy. The combination of machine learning models with lexicon-based features is a notable direction in recent studies, aiming to leverage the strengths of both approaches

Literature Review

• Jeelani Ahmed and Muqeem Ahmed, 6 Apr 2020

A Framework for Sentiment Analysis of Online News Articles

https://drive.google.com/file/d/10pe0lVqU78Ds95ms6TvzDvc-eyzld4yw/view?usp=sharing

This document delves deeper into machine learning methodologies applied to sentiment analysis. It reviews the evolution from traditional lexicon-based methods to more advanced machine learning models:

• **Supervised Learning Models:** The literature review highlights the use of supervised learning models, such as Naive Bayes, Support Vector Machines (SVM), and Decision Trees, which have been widely adopted in sentiment analysis. These models are trained on labeled datasets to classify text into sentiment categories. Pang et al. (2002) were among the first to apply machine learning techniques to sentiment classification, showing significant improvement over rule-based methods.

• **Deep Learning Techniques**: The paper further explores the rise of deep learning in sentiment analysis. Neural networks, especially Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have shown promising results due to their ability to capture complex patterns and context in text data. The use of pretrained models like Word2Vec and GloVe for embedding words into vectors has also been instrumental in improving the performance of sentiment analysis models.

Literature Review

• **Hybrid Models:** Recent studies have proposed hybrid models that combine lexicon-based approaches with machine learning algorithms. These models aim to improve sentiment classification by utilizing lexicon-based sentiment features as input to machine learning models. This hybrid approach addresses some limitations of purely lexicon-based methods, such as handling negations and capturing contextual information more effectively.

• **Challenges and Future Directions:** The literature also discusses the ongoing challenges in sentiment analysis using machine learning, such as handling sarcasm, detecting implicit sentiment, and the need for large labeled datasets. Future research directions include the development of more robust models that can handle multilingual sentiment analysis and the integration of external knowledge sources to improve model performance.

• Prateek Majumder , 29 Nov, 2021

https://www.analyticsvidhya.com/blog/2021/11/web-scraping-a-news-article-and-performing-sentimentanalysis-using-nlp/

Goal of our Project

The goal of our project is to develop a comprehensive system that can automatically analyze and classify the sentiment of news articles related to Indian politics. By accurately predicting the sentiment, we aim to provide valuable insights into how different political parties are represented in the media and help everyone understand media bias and public perception.

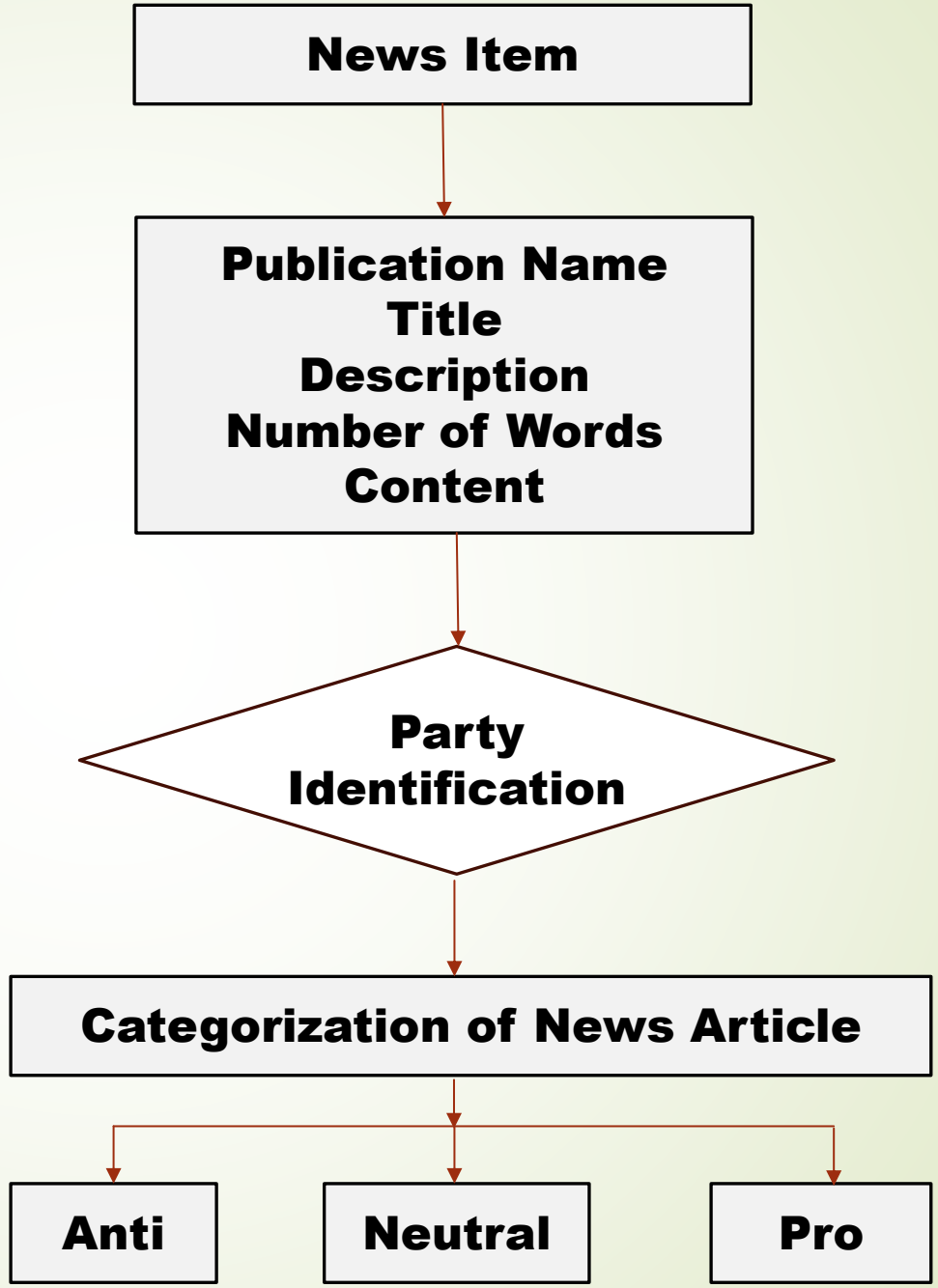
Objectives

- To extract relevant news articles focusing on Indian political news using the News API.
- To conduct sentiment analysis on the news content to classify it as positive, negative, or neutral.
- To identify the political party being discussed and analyze the sentiment towards it.
- To build a sentiment model by calculate sentiment scores using sentiment intensity analyzer (polarity scores)
- Predicting the overall sentiment of the news articles.
- Extracting result data into an excel file
- To provide insights into the portrayal of political parties in print media.

Polarity

- The polarity score ranges from -1 to 1.
- A score of -1 means the words are super negative, like “disgusting” or “awful.”
- A score of 1 means the words are super positive, like “excellent” or “best.”

Project Pipeline



Data Extraction

We use a News API key to extract relevant news articles from various sources focusing on Indian politics.

apiKey = "f57a815c8fdb43acacc42aa1f1b814d8"

Web Scraping

Utilizing the Beautiful Soup library in a Colab notebook, we scrape the content of the extracted articles to gather the text for analysis.

Sentiment Analysis

We apply Natural Language Processing (NLP) techniques to analyze the sentiment of the news content, categorizing it as positive, negative, or neutral.

Party Identification

The system identifies the political party being discussed in each article and correlates the sentiment with the party.

Sentiment Scoring

We calculate sentiment scores based on the analysis and use these scores to predict the overall sentiment of the news article.

Prediction and Insights

The results are compiled to provide insights into the sentiment trends in print media, focusing on political coverage.

Evaluation

- Polarity score >= 0.5

  Predicts sentiment as positive

- Polarity score <= - 0.5

  Predicts sentiment as Negative

- -0.5 < Polarity score < 0.5

  Predicts sentiment as Neutral

What we had done till now?

- Extracted a significant number of news articles related to Indian politics using the News API.
- Successfully implemented web scraping techniques using the Beautiful Soup library to gather the necessary content from the articles.
- performed sentiment analysis using NLP on the extracted news content.
- Started analyzing the sentiment towards different political parties and compiling the results by using polarity scores.
- Extracting results in an excel file for use.

Output

Total 214 News Articles are extracted from the News API Key and sentiment is predicted for all these 214 articles.

The preview of the output file is shown in the next slide.

Link for the output Excel file

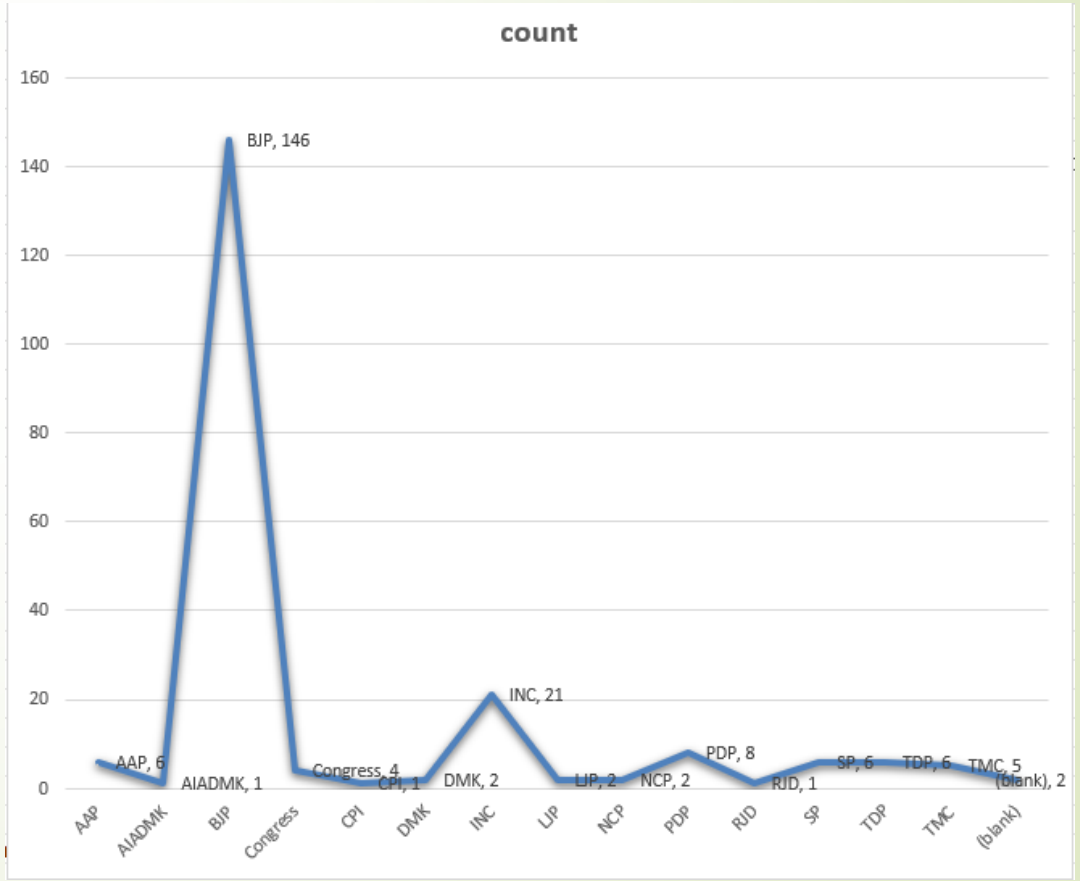
https://docs.google.com/spreadsheets/d/1fQdykYAOtT6k19Z1HsWLcxsuU50zKNl6/edit?usp=drive_link

	A	B	C	D	E	F	G	H	I	J	K
1	source	author	title	description	url	urlToImage	publishedAt	full_content	sentiment	Senti	party
2	{'id': 'al-ja	Al Jazeera Staff	India electior	Absence of the	https://www.aljazeera.com/news/202	https://www.aljazeera.co	2024-05-13T04:56:59Z	The absence of the	0.9631	Positive	PDP
3	{'id': 'al-ja	Al Jazeera Staff	India Lok Sab	People from acr	https://www.aljazeera.com/news/202	https://www.aljazeera.co	2024-05-12T05:12:01Z	A total of 96	0.5673	Positive	BJP
4	{'id': 'rt', 'r	RT	Indian capita	Delhi goes to th	https://www.rt.com/india/598233-dell	https://mf.b37mrtl.ru/file	2024-05-25T11:51:47Z	The sixth phase of	0.9449	Positive	BJP
5	{'id': 'rt', 'r	RT	Trans Democ	Non-binaries ex	https://www.rt.com/india/597965-five	https://mf.b37mrtl.ru/file	2024-05-22T03:33:39Z	In the ongoing seven-	0.9947	Positive	BJP
6	{'id': None	PTI	Lok Sabha ele	Voting begins in	https://www.thehindubusinessline.cor	https://bl-i.thgim.com/pu	2024-05-25T03:41:43Z	Get businessline apps	0.9685	Positive	BJP
7	{'id': None	Arvin Alaigh	Indiaâ€™s Ele	Arvin Alaigh	https://www.thenation.com/article/ar	https://www.thenation.c	2024-06-10T14:31:54Z		0.9985	Positive	BJP
8	{'id': None	mint	INDIA was a g	INDIA was a we	https://www.livemint.com/opinion/co	https://www.livemint.cor	2024-06-04T15:45:07Z		0.9987	Positive	INC
9	{'id': 'the-t	ET Online	Lok Sabha ele	The outcome of	https://economictimes.indiatimes.com	https://img.etimg.com/th	2024-05-25T05:27:08Z	Modi 3.0 Live	0.9903	Positive	BJP
10	{'id': None	BL New Delhi Bur	Lok Sabha Ele	India Elections F	https://www.thehindubusinessline.cor	https://bl-i.thgim.com/pu	2024-06-03T16:53:52Z	Get businessline apps	0.9943	Positive	BJP
11	{'id': 'the-t	PTI	Delhi Lok Sab	Somnath Bharti	https://economictimes.indiatimes.com	https://img.etimg.com/th	2024-05-24T08:47:37Z	Somnath Bharti is a	0.93	Positive	AAP
12	{'id': None	Andrew Deck	Indian journa	In October 2023	https://www.niemanlab.org/2024/05/i	https://www.niemanlab.c	2024-05-22T19:35:24Z	In October 2023,	0.9988	Positive	BJP
13	{'id': 'the-t	ET Online	Congress-AAI	Over 1.5 crore v	https://economictimes.indiatimes.com	https://img.etimg.com/th	2024-05-25T04:55:42Z	Modi 3.0 Live	0.9879	Positive	BJP
14	{'id': None	The Hills Times	Lok Sabha, 2(	NEW DELHI, Ma	https://thehillstimes.in/national/lok-sabha-2025-assembly-polls-		2024-05-18T20:25:16Z	11 June, 2024 |	0.9972	Positive	LJP
15	{'id': None	K V Kurmanath	Lok Sabha Ele	N Chandrababu	https://www.thehindubusinessline.cor	https://bl-i.thgim.com/pu	2024-06-04T08:13:16Z	Get businessline apps	0.9953	Positive	TDP
16	{'id': 'the-t	PTI	Successful J-k	On December 1	https://economictimes.indiatimes.com	https://img.etimg.com/th	2024-05-26T07:02:17Z	Modi 3.0 Live	0.4857	Neutral	BJP
17	{'id': 'the-t	ET Online	Lok Sabha ele	The political cor	https://economictimes.indiatimes.com	https://img.etimg.com/th	2024-05-24T14:01:10Z	Modi 3.0 Live	0.962	Positive	BJP
18	{'id': 'the-t	ET Online	Lok Sabha Ele	India General El	https://economictimes.indiatimes.com	https://img.etimg.com/th	2024-05-25T04:22:07Z	Modi 3.0 Live	0.9884	Positive	BJP
19	{'id': None, 'name': 'Busines		Letters to the	Letters to Editor	https://www.thehindubusinessline.cor	https://www.thehindubu	2024-06-04T16:55:30Z	Get businessline apps	0.9972	Positive	BJP
20	{'id': 'the-t	ET Bureau	Congress-AAI	Delhi has been a	https://economictimes.indiatimes.com	https://img.etimg.com/th	2024-05-23T18:11:40Z	Modi 3.0 Live	0.9626	Positive	BJP
21	{'id': 'the-t	ET Online	Election Resu	Early trends in t	https://economictimes.indiatimes.com	https://img.etimg.com/th	2024-06-04T13:40:48Z	Modi 3.0 Live	0.952	Positive	BJP
22	{'id': None	Neville Lazarus	India PM's pa	On his campaigi	https://news.sky.com/story/india-elec	https://e3.365dm.com/24	2024-05-18T13:23:00Z	Prime Minister	0.9806	Positive	PDP
23	{'id': None	The Hills Times	Constitution	NEW DELHI, Jun	https://thehillstimes.in/national/constitution-our-guiding-light-w		2024-06-04T19:15:12Z	11 June, 2024 |	0.9965	Positive	INC
24	{'id': None	The Hills Times	TDP pledges t	NEW DELHI, Jun	https://thehillstimes.in/national/tdp-pledges-to-revive-shelved-a		2024-06-04T20:55:10Z	11 June, 2024 |	0.9485	Positive	TDP

Analysis

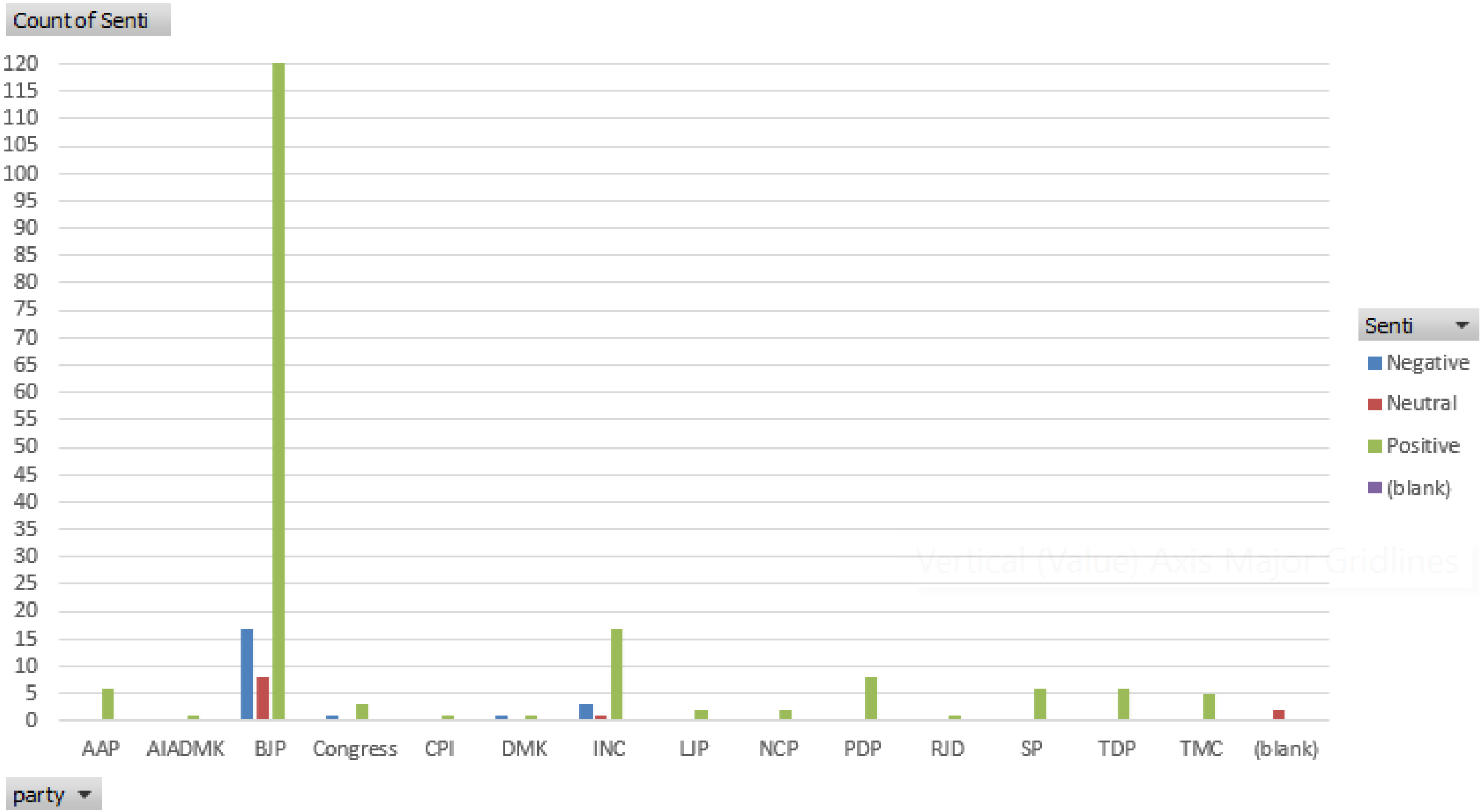
There are total 213 articles extracted and the BJP part has the maximum count of articles.

party	count
AAP	6
AIADMK	1
BJP	146
Congress	4
CPI	1
DMK	2
INC	21
LJP	2
NCP	2
PDP	8
RJD	1
SP	6
TDP	6
TMC	5
(blank)	2
total	213



By the analysis of all these articles we have total 213 articles in which 180 are positive, 22 are negative and 11 articles are neutral.

Count of Senti	Column Labels				
Row Labels	Negative	Neutral	Positive	(blank)	Grand Total
AAP			6		6
AIADMK			1		1
BJP	17	8	121		146
Congress	1		3		4
CPI			1		1
DMK	1		1		2
INC	3	1	17		21
LJP			2		2
NCP			2		2
PDP			8		8
RJD			1		1
SP			6		6
TDP			6		6
TMC			5		5
(blank)		2			2
Grand Total	**22**	**11**	**180**		**213**


Count of Senti
120
115
110
105
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
AAP
AIADMK
BJP
Congress
CPI
DMK
INC
LJP
NCP
PDP
RJD
SP
TDP
TMC
(blank)
party
Senti
Negative
Neutral
Positive
(blank)
Vertical (Value) Axis Major Gridlines

References

- Antony Samuels, John Mcgonical 29 Nov, 2021

News Sentiment Analysis

https://drive.google.com/file/d/1-OQtyrlj5G2UTjVTWI_LvM4vClag3-7n/view?usp=drivesdk

- Jeelani Ahmed and Muqeem Ahmed , 6 Apr 2020

A Framework for Sentiment Analysis of Online News Articles

https://drive.google.com/file/d/10pe0lVqU78Ds95ms6TvzDvc-eyzld4yw/view?usp=sharing

- Prateek Majumder , 29 Nov, 2021

https://www.analyticsvidhya.com/blog/2021/11/web-scraping-a-news-article-and-performing-sentiment-analysis-using-nlp/

Acknowledgements

I am highly grateful to Sabudh Foundation, Mohali, for providing an opportunity to carry out twelve months of training from July 2023 - August 2024.

Dr. Shafila Bansal and Mr. Sanchit Umate have provided great help in carrying out my work and is acknowledged with reverential thanks. Without wise counsel and able guidance, it would have been impossible to complete the training in this manner. I would also like to thank the entire team of the Sabudh Foundation.

I would also thank my friends who devoted their valuable time and helped me in all possible ways towards the successful completion of this training.

Thank You