GWU CSCI 6443 Data Mining

Instructor-Prof. Paul Melby

Author-Group 11: Snigdha Nigam

December 15, 2025

Sentiment Analysis of Media Coverage of the 2025 Nepal Prime Ministerial Election.

Abstract-

In September 2025, Nepal gained global attention after electing its first female Prime Minister using Discord as the "maiden". Unlike prior regime changes, this election was significantly shaped by online discourse and digital platforms, an unprecedented intersection of democracy and the internet. This paper is a discussion of sentiment analysis techniques used by me to examine how major English language media outliers portrayed this political transition. The database for this task was created by scraping web news articles and blogs. Experimental analysis was performed using a transformer, however, the final analysis was done using TextBlob and VADER. This project helped me learn the methodological differences between two sentiment analysis techniques previously mentioned. But more importantly it helped me get an intuitive understanding of two important concepts 1.**Overfitting vs Generalization**

and **2. Importance of collecting good quality data.** A very interesting thing that I realized about my code is that every time the scrapping script is run, it collects data from the most recent version of each website. This causes slight variations in the collected data which also affects the end results which can be observed through the difference in visualizations gathered on the presentation day- December 4th and the paper submission day- December 15th. I will talk about this discrepancy more in the Results section.

TextBlob provided a sentiment distribution for long form news articles which was quite different from the results produced by VADER which exhibited greater sensitivity to contextual modifiers.

Keywords-Nepal, Election, Prime Minister, Sentiment Analysis,Web-scraping, Database Transformer, TextBlob, VADER

Business Problem

The problem statement can be summed up as " How did the international media respond to the political transition of power in Nepal?"

Understanding media sentiment around this issue has practical implications for political strategy. For stakeholders like political analysts, campaign managers, journalists, or data analysis firms, understanding how such events are framed in the media is essential in political decisions.

Data Collection-

I began by manually searching for websites related to the Nepal election held in September 2025. Searching for websites that were both in English language and not behind a paywall was a tedious task.

Data Sources-

1. https://www.wired.com/story/nepal-discord-gen-z-protests-vote-prime-minister-election

2. ‘More egalitarian’: How Nepal’s Gen Z used gaming app Discord to pick PM | Social Media News | Al Jazeera
3. From Streets to Discord: How Nepal’s Gen Z Toppled a Government | Carnegie Endowment for International Peace
4. How Nepal’s enraged Gen Z turned Discord into a political arena
5. Nepal’s Social Media Elects Their Prime Minister Through Discord – The New Dealer
6. Nepal’s Discord Revolution and the Press - Nieman Reports
7. Sushila Karki: How Gen Z protestors chose Nepal’s first woman prime minister on Discord | The Independent
8. Youth Group in Nepal Overthrows Government, Elects New Leader via Discord Poll
9. Sushila Karki - Wikipedia
10. Former chief justice sworn in as Nepal’s interim prime minister following deadly protests | CNN
11. The New Prime Minister of Nepal Was Elected on Discord – Joshua Foust
12. Democracy in the Age of Discord: Nepal’s Gen-Z chooses leader through social media - Constitutional Discourse
13. With Nepal in Flames, Gen Z Elects The New Prime Minister On Discord
14. Nepal's Gen Z chooses Discord to 'elect' interim PM: What is this app and how it works - The Times of India
15. Nepal Gen Zs endorse interim prime minister via Discord vote - Cryptopolitan
16. Youth Group in Nepal Overthrows Government, Elects New Leader via Discord Poll
17. Nepal’s Gen Z Elects New Interim Prime Minister Sushila Karki Via Discord Voting- Viral Memes- ‘New Era Of Elections’ | Viral - Times Now
18. Discord becomes polling booth as Nepal's Gen Z endorse new government
19. Nepal picks a new prime minister on a discord server days after social media ban | Hacker News
20. Nepal - Himalayan, Monarchy, Federal Republic | Britannica
21. Sushila Karki Becomes Nepal’s First Woman PM as Nation Creates History, Elects Leader Through Discord - The Logical Indian
22. Nepal’s PM Elected via Discord: A New Form of Democracy Emerging – Techquity India

23. Nepal's Gen Z protesters choose Discord to pick next PM. Here's everything about the chat app | Today News

Web Scraping Techniques-

Newspaper-

Newspaper3k is a Python library used for scraping news articles from online news articles. It automatically downloads an article from a given URL and extracts clean text, headlines, authors, publish dates, and images. Newspaper3k also scrapes information from hyperlinks, ads, user comments and other clickable items from the website, creating a very detailed yet noisy database. It helped scrap 402 articles.

Beautifulsoup-

Beautiful Soup is a Python library used for web scraping that helps extract information from HTML and XML pages. It works by parsing through the structure of a webpage and easily finding and collecting elements like headlines, links, paragraphs, or tables using simple commands. It help scrape 23 articles.

Cosine Similarity-

Cosine similarity helped create the "final dataset" with 12 articles that were most similar or relevant to the topic. A threshold of 0.45 was chosen and set by trial and error.

Methodology-

1. Making a list of weblinks that I wanted to scrape	2. Performing scraping with newspaper and beautifulsoup	3. Performing cosine similarity between the articles base created by both scraping techniques	4. Post similarity check data cleaning (removing duplicates, filling empty columns)
5.Sentiment analysis- Textblob	6. VADER	7.Visualisation	8. Decision(which is better)

I started by first creating a list of newspaper article URLs to be scraped. Then built two datasets using web-scraping tools, Newspaper3k and Beautiful Soup. Next, calculated the cosine similarity between the articles obtained from both methods to extract the articles relevant to Nepal's election. After this comparison, the data is cleaned by removing duplicates and handling missing values. Sentiment analysis is then performed on the cleaned articles using TextBlob and VADER.

Sentiment Analysis-

Example 1: showing three articles- TextBlob=Neutral, VADER=Negative

title
The Inside Story of How Gen Z Toppled Nepal's Leader and Chose a New One on Discord
Make Your Day
At least 19 killed in Nepal protests over social media ban, corruption

1. TextBlob

TextBlob is a lexicon-based natural language processing library in Python that provides a simple approach to sentiment analysis. It is particularly well-suited for long-form textual data such as news articles. It computes sentiment using two primary metrics: **polarity**, which ranges from –1 (negative) to +1 (positive), and **subjectivity**, which ranges from 0 (objective) to 1 (subjective). TextBlob determines these scores by assigning predefined sentiment values to individual words and averaging them across the entire text. This aggregation makes it effective for capturing the dominant emotional tone of full-length articles rather than overemphasizing isolated phrases.

polarity	subjectivity	sentiment
0.07487460815047020	0.4182027168234070	neutral
0.0	0.0	neutral
-0.12281612420501300	0.3610035540591100	negative

2. VADER

VADER stands for Valence Aware Dictionary and Sentiment Reasoner is a rule-based sentiment analysis tool designed to capture sentiment intensity in text by leveraging a predefined lexicon of words scored for positive and negative valence. Along with individual word sentiment, VADER accounts for linguistic features such as emojis, degree modifiers, punctuation, capitalization, and contrastive conjunctions, enabling it to model contextual sentiment more effectively than simple bag-of-words approaches. VADER produces a compound sentiment score normalized between –1 and +1, along with separate positive, negative, and

neutral proportions. While it performs particularly well on short, informal text such as social media posts, its application to long-form news articles in this project resulted in more compressed sentiment scores, as positive and negative terms often offset each other across lengthy narratives.

sentiment	vader_scores	vader_sentiment
neutral	{'neg': 0.134, 'neu': 0.763, 'pos': 0.103, 'compound': -0.9446}	negative
neutral	{'neg': 0.0, 'neu': 0.0, 'pos': 0.0, 'compound': 0.0}	neutral
negative	{'neg': 0.176, 'neu': 0.77, 'pos': 0.054, 'compound': -0.9987}	negative

For the first article titled-"The Inside Story of How Gen Z Toppled Nepal's Leader and Chose a New One on Discord" the VADER sentiment is negative while the TextBlob sentiment is neutral.

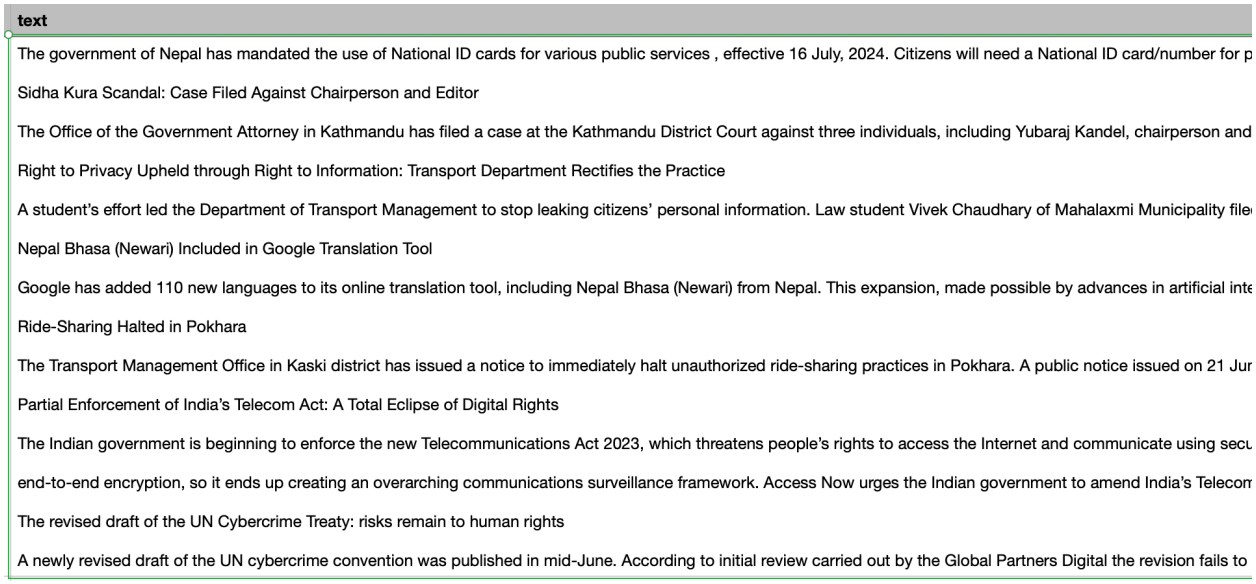
VADER gives out word scores based on very specific hard wired rules.

A statement from the article-"**It ended with protests, violence, and an online poll to pick the new prime minister.**"

- VADER analyzed and flagged words like **end, protest, violence, and new** as an overwhelmingly negative sentiment.
- **Textblob flagged the above sentence as neutral because it is just a 'statement of fact'.**
- **This insight comes from combining both Polarity (independent meaning- positive, negative or neutral) and their subjective meaning.**

- **Lexicon analysis as well as subjectivity. Polarity is weighted across a sentence.**

Example 2 : TextBlob=Neutral, VADER-Negative



text
The government of Nepal has mandated the use of National ID cards for various public services , effective 16 July, 2024. Citizens will need a National ID card/number for p
Sidha Kura Scandal: Case Filed Against Chairperson and Editor
The Office of the Government Attorney in Kathmandu has filed a case at the Kathmandu District Court against three individuals, including Yubaraj Kandel, chairperson and
Right to Privacy Upheld through Right to Information: Transport Department Rectifies the Practice
A student's effort led the Department of Transport Management to stop leaking citizens' personal information. Law student Vivek Chaudhary of Mahalaxmi Municipality file
Nepal Bhasa (Newari) Included in Google Translation Tool
Google has added 110 new languages to its online translation tool, including Nepal Bhasa (Newari) from Nepal. This expansion, made possible by advances in artificial inte
Ride-Sharing Halted in Pokhara
The Transport Management Office in Kaski district has issued a notice to immediately halt unauthorized ride-sharing practices in Pokhara. A public notice issued on 21 Jur
Partial Enforcement of India's Telecom Act: A Total Eclipse of Digital Rights
The Indian government is beginning to enforce the new Telecommunications Act 2023, which threatens people's rights to access the Internet and communicate using secu
end-to-end encryption, so it ends up creating an overarching communications surveillance framework. Access Now urges the Indian government to amend India's Telecom
The revised draft of the UN Cybercrime Treaty: risks remain to human rights
A newly revised draft of the UN cybercrime convention was published in mid-June. According to initial review carried out by the Global Partners Digital the revision fails to

This text talks about mandated use of National ID cards for public service, right to privacy through right to information, unauthorized Ride Sharing Halted in Pokhara and a Telecommunications Act 2023 issued by the Indian Government. Going through this article manually made it clear that these are just "facts" stated in light of the recent election. **Once again we see TextBob correctly identifying facts as neutral. While VADER flags the lack of "negative" words as a net "positive" sentiment.**

TextBlob- NEUTRAL

polarity	subjectivity	sentiment
0.033293249959916 60	0.36195486612153 30	neutral

VADER-POSITIVE

vader_scores	vader_sentiment
{'neg': 0.056, 'neu': 0.868, 'pos': 0.076, 'compound': 0.8768}	positive

Example 3- VADER and TextBlob Agree!

text
At least 19 people have been killed and dozens injured during violent protests against the government's social media ban and alleged corruption in Nepal, according to aut
On Monday, some protesters forced their way into the Parliament complex in the capital, Kathmandu, by breaking through a barricade, a local official said.
One protester told the ANI news agency that the police had been firing "indiscriminately".
"[They] fired bullets which missed me but hit a friend who was standing behind me. He was hit in the hand," the protester said.
Seven people died at the National Trauma Centre, chief medical superintendent Dr Badri Rijal told The Associated Press news agency.
"Many of them are in serious condition and appear to have been shot in the head and chest," Rijal said.
Families waited anxiously outside for news of their relatives while people gathered to donate blood.
Police officer Shekhar Khanal told Reuters that more than 100 people, including 28 police personnel, were receiving medical treatment for their injuries. Two people were ki
The United Nations has called for accountability following the bloodshed. "We are shocked by the killings and injury of protesters in Nepal today and urge a prompt and tra
Advertisement
Anurag Acharya, director at the Kathmandu-based think tank Policy Entrepreneurs Inc, says the trigger for the protests may have been the social media ban, but the underl
"This is the generation that grew up with the optimism and dream of an inclusive new Nepal, a dream that remains elusive a decade after the new constitution was promulg
"The reality for today's Gen Z is that there are few livelihood prospects at home, which forces thousands to migrate abroad for studies and jobs. So, the frustration with uns
'Youths against corruption'
Thousands of young people, including students in their school and college uniforms, joined the protest, holding signs that read "Shut down corruption and not social media
Ikshama Tumrok, a 20-year-old student, told the AFP news agency that she was protesting against the "authoritarian attitude" of the government.
"We want to see change. Others have endured this, but it has to end with our generation," she said.
Last week, the government decided to block access to several social media platforms, including Facebook, YouTube and X, fuelling anger among young Nepalis.
According to officials, the decision was taken because platforms had failed to register with authorities in a crackdown on misuse, including fake social media accounts use
Organisers of the protests, which have been dubbed "demonstrations by Gen Z", have said their civil disobedience reflects how the youth feel about the government decis
One protester told ANI that this was "the protest by the new generation in Nepal".
Muktiram Rijal, a spokesperson for the Kathmandu District Administration Office, told the Reuters news agency that the police had orders to use water cannon, batons and

The text talks about deaths and injuries resulting from several violent protests against Nepal's social media ban a few days before the Election.

Contains words like- killed, forced, indiscriminately, shot, Trauma Center, Died, shocked, injury etc.

VADER- used its pre-defined labels to correctly flag this piece of text as "negative".

vader_scores	vader_sentiment
{'neg': 0.176, 'neu': 0.77, 'pos': 0.054, 'compound': -0.9987}	negative

Both lexical analysis and subjective analysis indicates that this text is "negative".

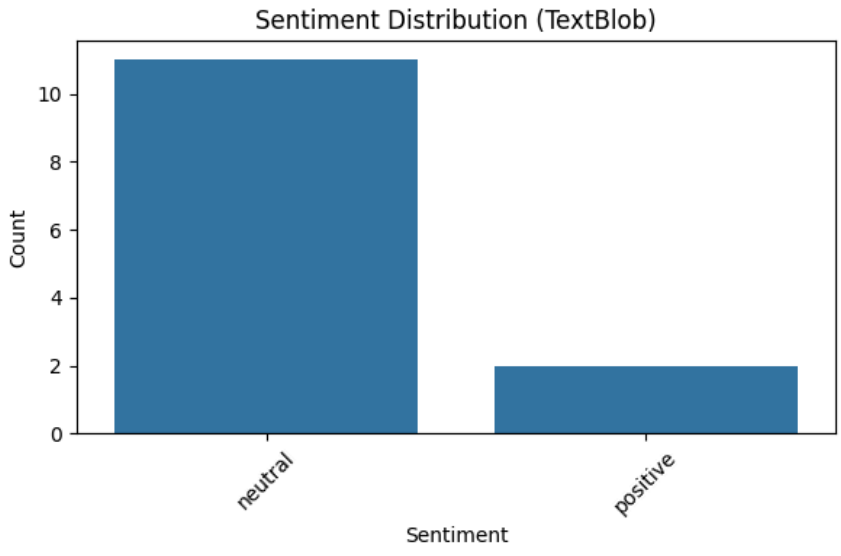
polarity	subjectivity	sentiment
-0.12281612420501300	0.3610035540591100	negative

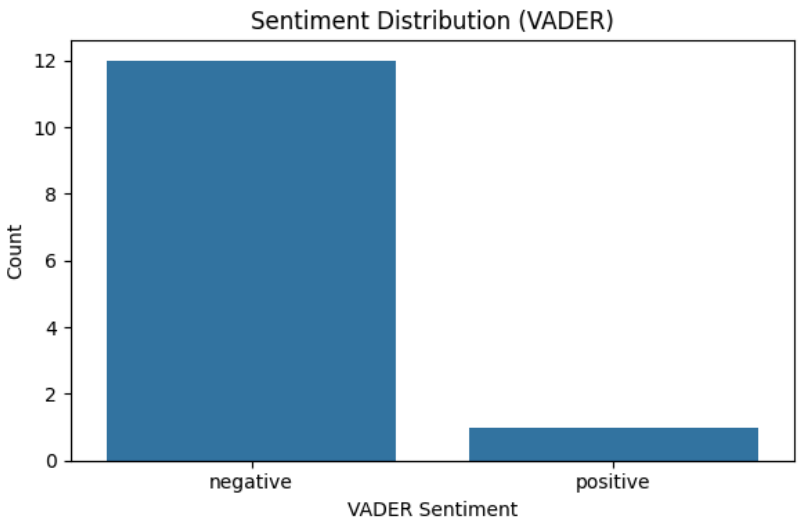
Personal Observation-

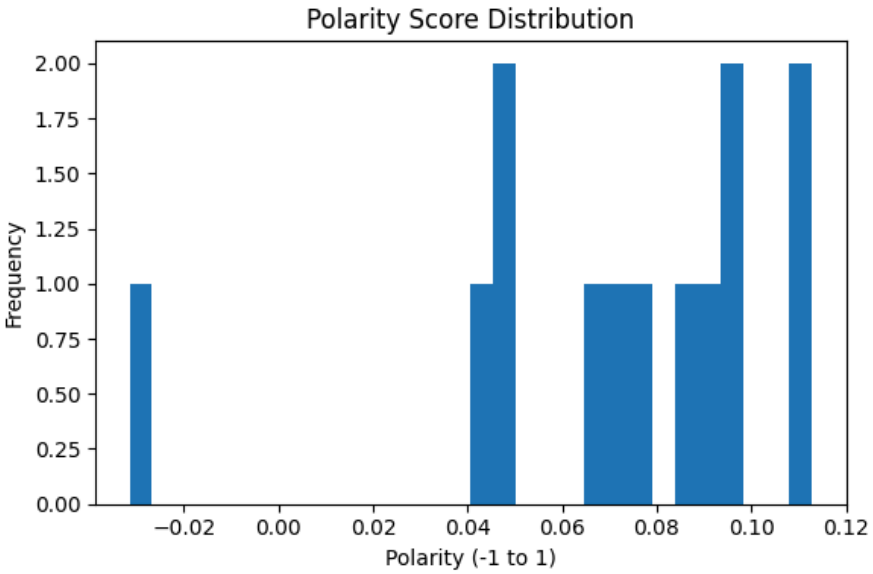
For long form content- VADER overfits! while Textblob generalizes a tad.

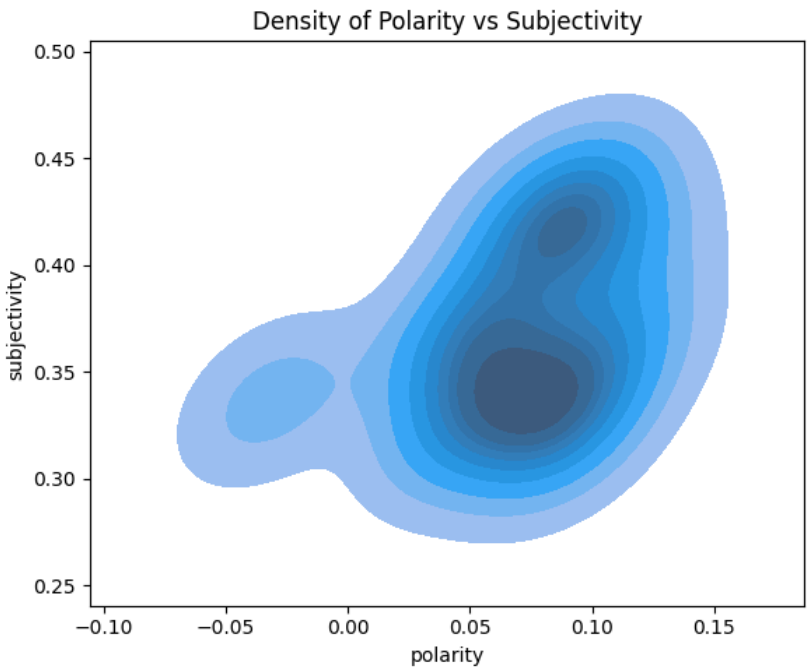
This ties back to lecture discussion about model performance. VADER overfits because it is a complex model with a very small sample size of training labels(about 900 words pre defined in its library).

Results- from the scraped dataset on December 4th

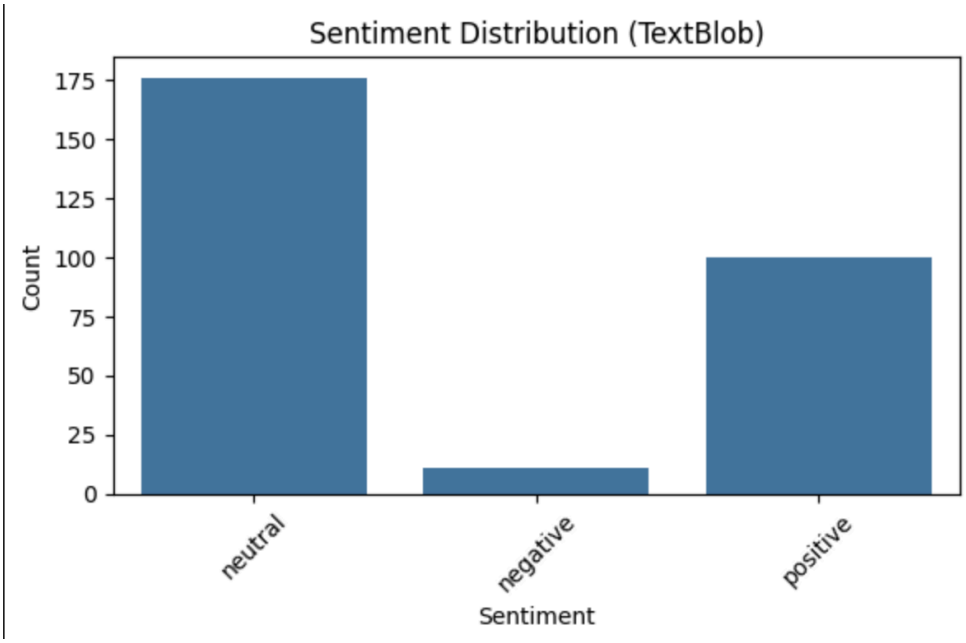


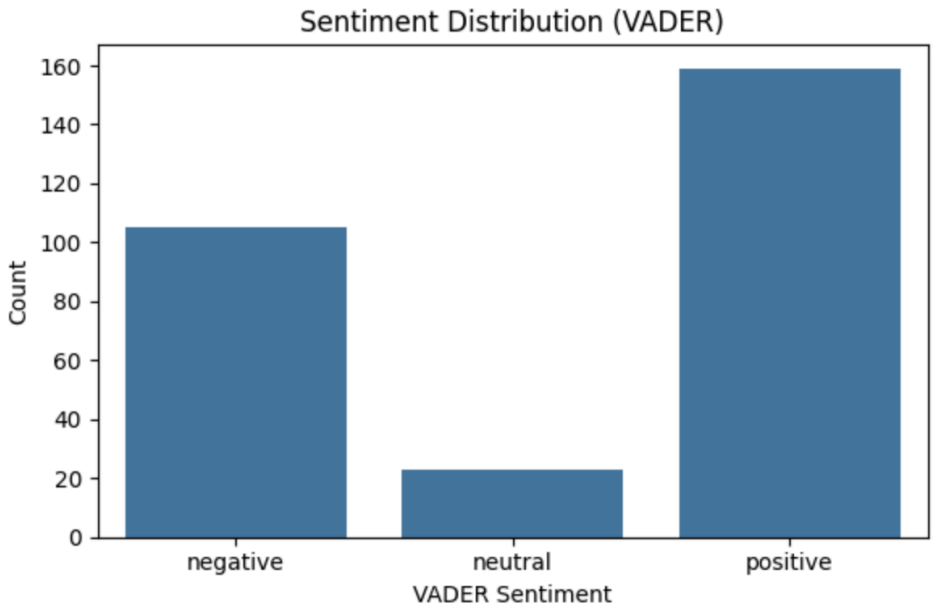

Sentiment Distribution (VADER)
Count
12
10
8
6
4
2
0
negative
positive
VADER Sentiment

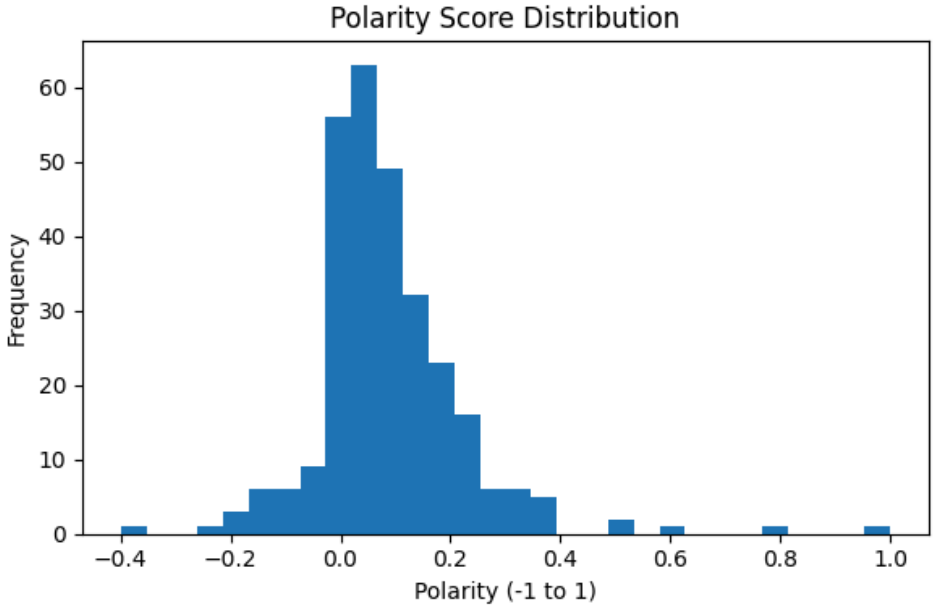

Polarity Score Distribution
Frequency
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.02
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Polarity (-1 to 1)

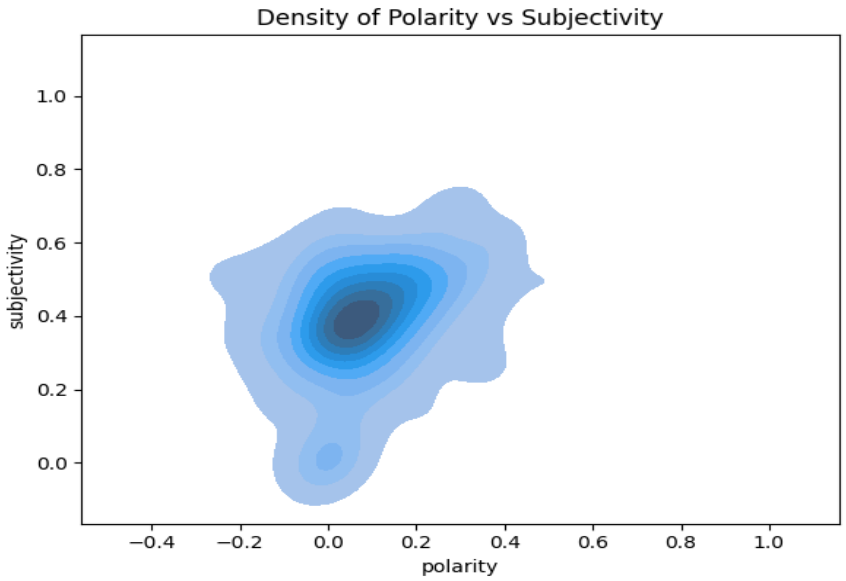


Results- From the scraped dataset on December 15th




Sentiment Distribution (VADER)
Count
0
20
40
60
80
100
120
140
160
negative
neutral
positive
VADER Sentiment


Polarity Score Distribution
Frequency
0
10
20
30
40
50
60
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
Polarity (-1 to 1)



Comparison between the two distinct results-

Based on data scrapped on Dec 15th	**Based on data scrapped on Dec 4th**
Polarity vs Subjectivity graph- Captures a broader spectrum(a wider range) of sentiment polarity, perhaps due to newer or more diverse scraped content. Indicates greater variation in opinionated vs factual language in the updated dataset. **The text is more positive and more subjective.**	**Polarity vs Subjectivity graph-** Shows a more compact and concentrated distribution, indicating greater variability in sentiment. **Text is less positive and less subjective.**
TextBLOB data analysis-	**TextBLOB data analysis-**

Shows all three sentiments (neutral, positive, and negative) with a much larger volume of data. It shows that Dec 15 data includes complaints, feedback, or problem reports. **Quality of data is varied which results in a thorough analysis.**	Shows only neutral and positive sentiments and no negative sentiment at all. While Dec 4 content is perhaps more informational or neutral. **There is less variance in the data collected on this day. Results are more skewed.**
VADER data analysis- There is a **large number of positive sentiment** data points, with relatively fewer negative and neutral sentiments. **VADER performed surface-level lexical analysis / gave overfitted results**	**VADER data analysis-** There is **a large number of negative sentiment** compared to the positive sentiment. **VADER performed surface-level lexical analysis / gave overfitted results**
Polarity Score Distribution- A bell shaped curve! **Data appears to be labelled in a gaussian fashion.** Shows a wider spread of polarity scores, ranging from clearly negative to strongly positive values.	**Polarity Score Distribution-** Tightly clustered around low, mildly positive scores with far fewer extreme values. Probably because the script captured a **smaller or more homogeneous snapshot of content**, resulting in a narrow polarity range.

Conclusion

It was interesting to see how data scraped from the same set of URLs but on different days introduced variations in the collected data and caused significant changes to the end result. It is also evident that the results gathered on December 15th show wider spread of polarity possibly because of the quality of data that was collected on this day.

The condition of the websites directly affect the quality and relevance of the scraped content. A robust dataset exposes the algorithm to diverse patterns which leads to uncovering of real trends in the data rather than skewed results or noisy results.

It is also abundantly clear that Textblob uses better linguistic understanding (lexicon analysis+subjectivity) to give a much more accurate and reliable output.

With analysis done on a wider range/ better quality of data I can say that I prefer the analysis done by Textblob-

1. Using polarity as a metric, Textblob provides a more factual analysis of the text collected from these articles.
2. VADER seems to sway its output with short bursts of emotional words **(protest='always negative', new prime minister=' always positive').** It misses a lot of nuance which is extremely necessary for accurate sentiment analysis. It over amplifies the **intensity often resulting in a poor analysis of long form content where relative meaning of phrases and the context of the situation are extremely important to consider.**
3. Textblob explicitly **models subjectivity** which helps clearly distinguish facts from opinions. It can work well with text which has a lot of **sentimental noise which could otherwise trick algorithms like VADER.**
4. TextBlob works well on full articles, like newspaper text, because it computes polarity based on the sum of all words' sentiment.

5. Gives a polarity between -1 and +1, which is easy to interpret.

References

1.https://textblob.readthedocs.io/en/dev/

2. https://hex.tech/templates/sentiment-analysis/vader-sentiment-analysis/

3.https://newspaper.readthedocs.io/en/latest/

4.https://beautiful-soup-4.readthedocs.io/en/latest/

5.cosine_similarity — scikit-learn 1.8.0 documentation