Embracing Public Sentiment: A New Dimension in Equity Portfolio Management

Akhil Nagabhyiru , Dhrithi Pradeep Alva, Surya Kalva, Uday

BIA 660 A - Team 5

Objectives

- Leverage public sentiment analysis to identify emerging trends and market shifts
- Detect early warnings about market changes and potential crises
- Augment technical analysis with granular sentiment analysis on each equity for a robust risk management strategy

Data Collection and Preprocessing

- Web scraping was employed to collect data related to many different companies from two different sources, i.e., the Hindenburg Research website and Yahoo News.
- To parse the HTML content of the website and extract the relevant information, the Beautiful Soup library was used.
- Once the data was collected, the text content was preprocessed by converting all the text to lowercase to avoid any discrepancies due to casing. Special characters and punctuation marks were removed to obtain a clean text corpus.
- Stopwords, i.e., commonly used words that do not carry much meaning, were filtered out using the NLTK library.

The below data scraped from Yahoo News for Adani and Nikola . Similarly we can check for other companies with our code.

	title	summary	source
0	Adani Wilmar Q4 results: Profit for Adani grou...	Adani Wilmar said its revenue from operations ...	Yahoo News
1	Adani Wilmar, Adani Enterprises, Adani Total, ...	Shares of Adani Total Gas have cracked 74.40 p...	Yahoo News
2	Adani Flagship's 138% Profit Jump to Aid Growt...	(Bloomberg) -- Adani Enterprises Ltd.'s latest...	Yahoo News
3	Adani Enterprises share price surges 1% today ...	Adani Enterprises shares have risen 13% in the...	Yahoo News
4	Adani power Q4 results: Net profit rises 12.9 ...	Adani Power has reported a 12.9 percent rise i...	Yahoo News

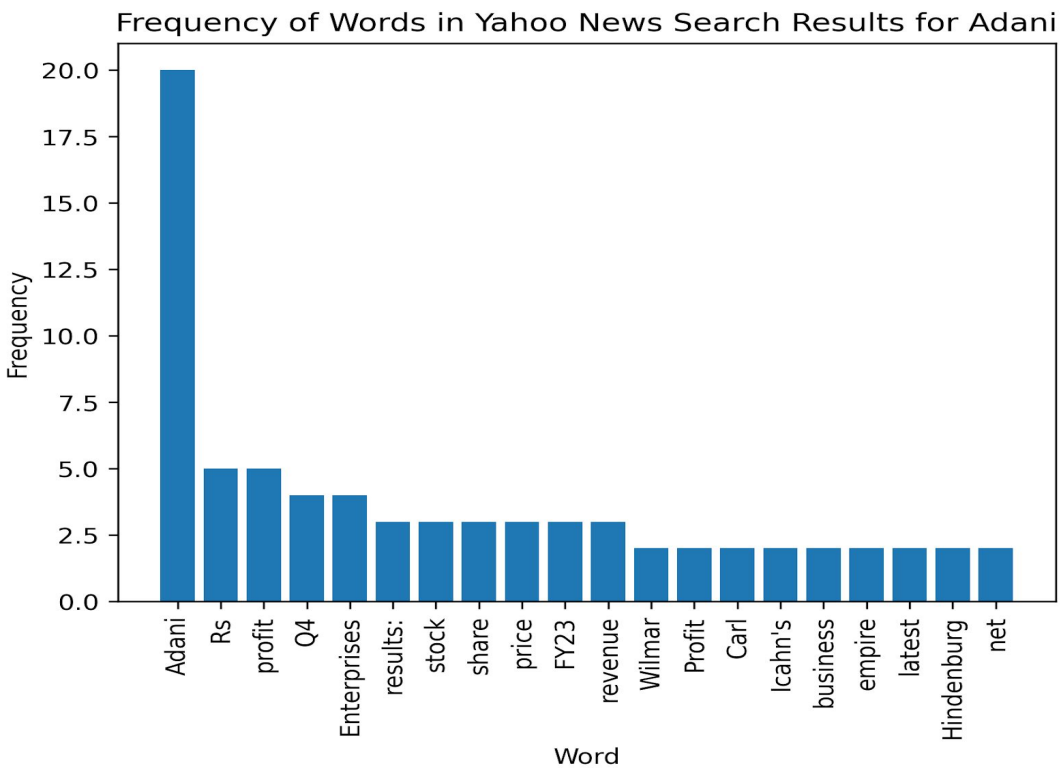
	title	summary	source
0	NIKOLA AND VOLTERA ENTER INTO A DEFINITIVE STR...	Nikola Corporation (Nasdaq: NKLA), a global le...	Yahoo News
1	Carl Icahn's business empire just became Hinde...	Here are some of the short seller's biggest be...	Yahoo News
2	WATTEV TO TAKE DELIVERY OF FIRST BATCH OF 14 N...	Nikola Corporation (NASDAQ: NKLA), a global le...	Yahoo News
3	Nikola (NKLA) to Report Q1 Earnings: What's in...	The Zacks Consensus Estimate for Nikola's (NKL...	Yahoo News
4	Tom's Truck Center Adds Nikola Class 8 Tre Sem...	Tom's Truck Center, a commercial truck sales a...	Yahoo News

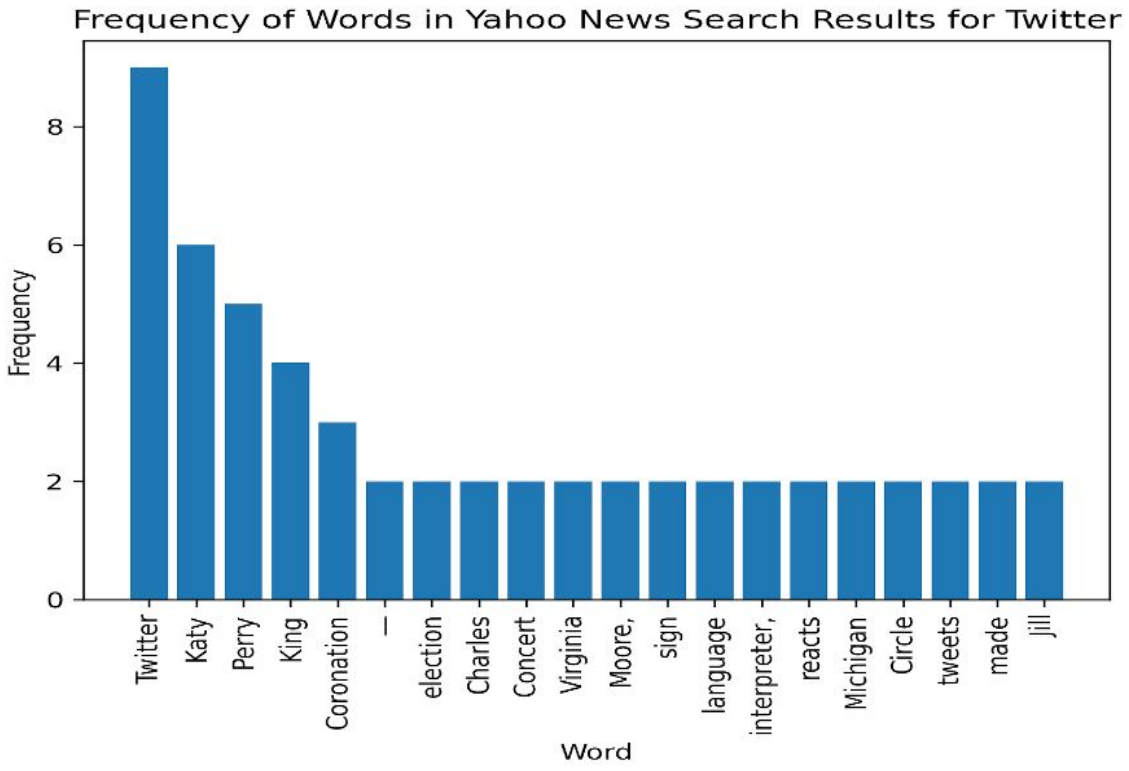
EDA (Exploratory Data Analysis:)

In this particular project, the EDA focused on several aspects related to the text data collected from the Hindenburg Research website and Yahoo News:

- Identifying the most common words within the text.
- Visualizing the distribution of words using word clouds.
- Performing sentiment analysis on the text.
- Most Common Words

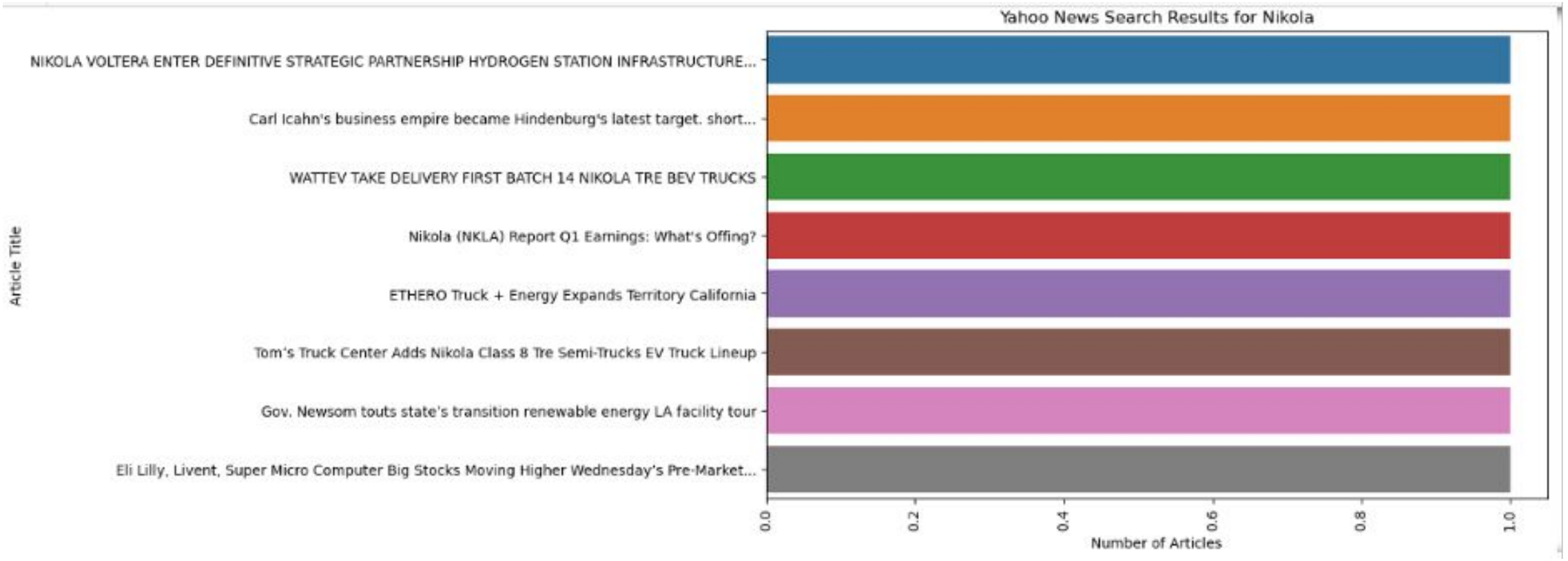
EDA-Frequency of Words In Yahoo News Search For Adani and Twitter





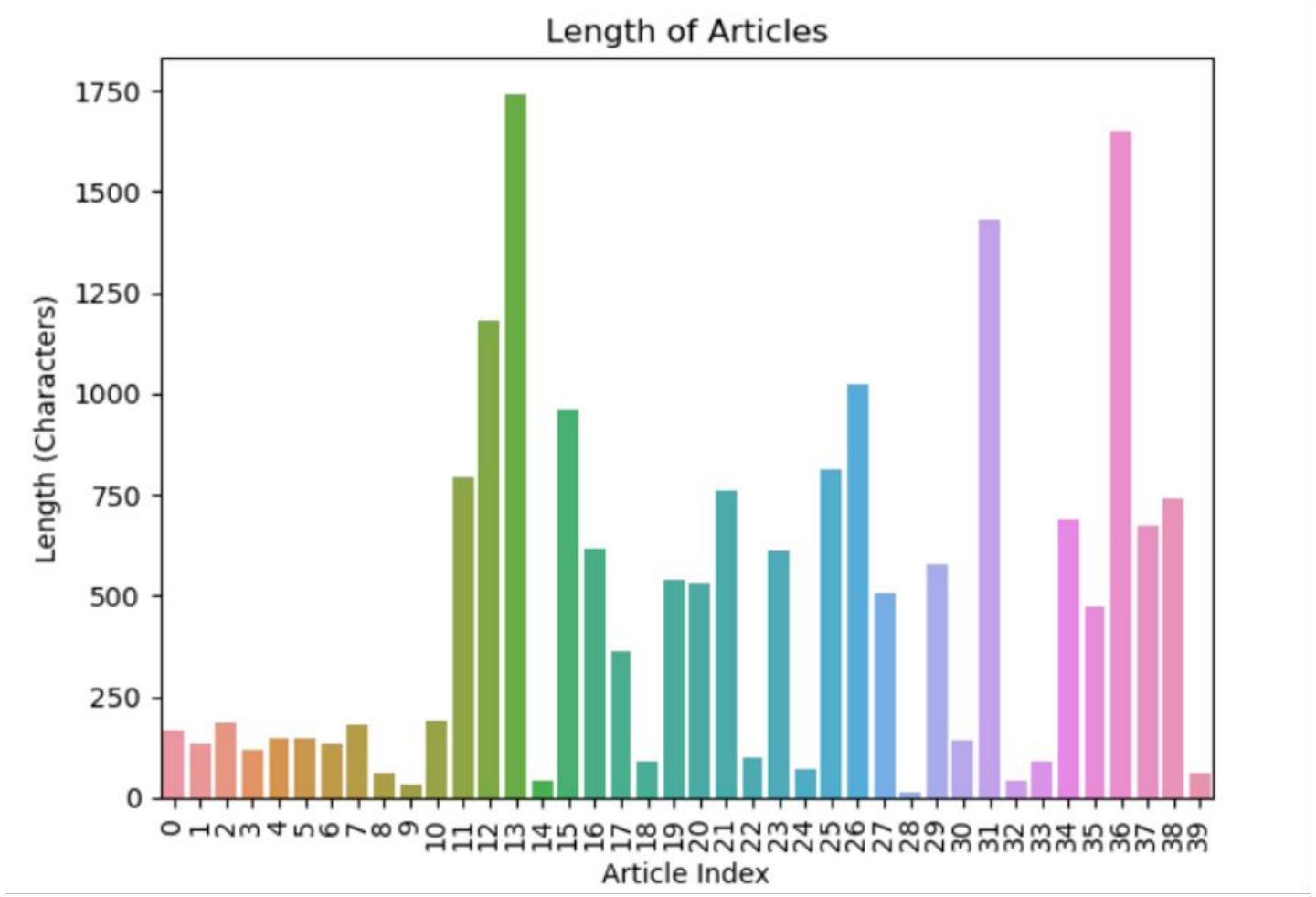
EDA

Our code scrapes for news articles related to companies of our interest, removes stop words from the titles and summaries, and creates a countplot of the article titles for each company. The resulting plot is displayed for each company. The countplot displays the number of articles for each unique title in the scraped data.

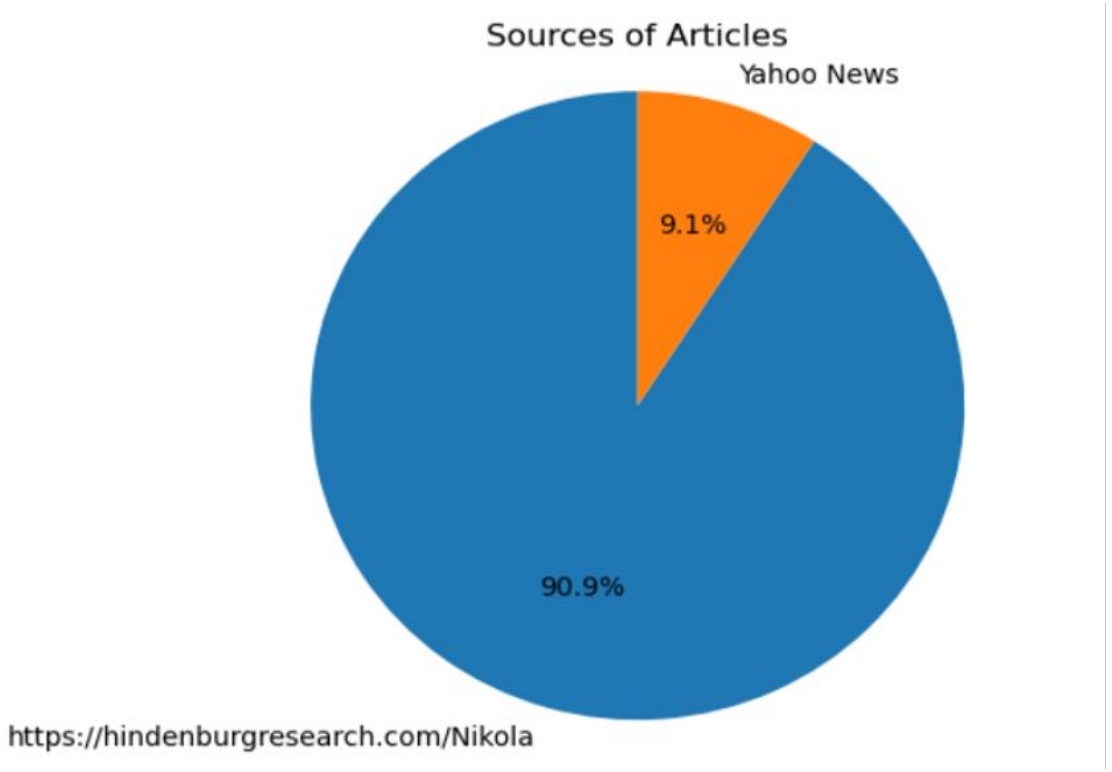


Review Analysis

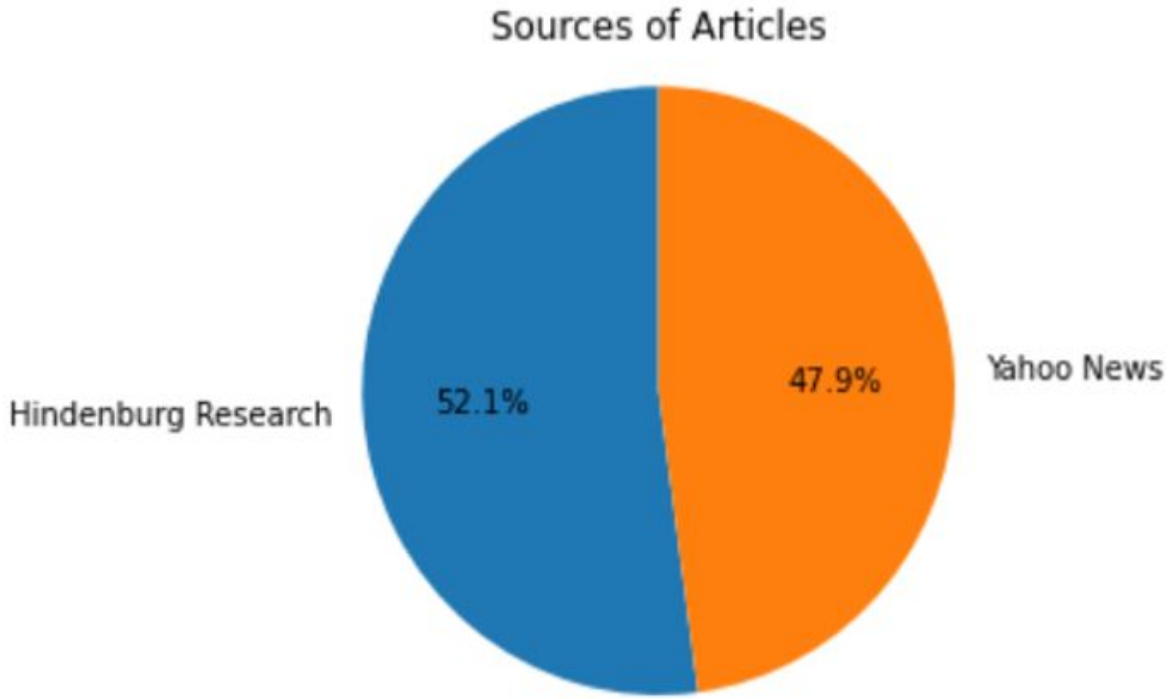
It is observed that, Average length of articles: 1149.841



- The distribution of sources for the articles in the dataset is visualized using a pie chart.
- The pie chart displays the percentage of articles from each source, and the labels show the name of the sources.
- The `autopct` parameter formats the percentages to one decimal place. The `axis('equal')` line ensures that the pie chart is drawn as a circle.
- This Image is for a Company Nikola



Here is another pie chart that visualizes the proportion of articles from different sources in a given dataset. It uses the value_counts function to count the number of articles from each source, then creates a list of labels and sizes from the resulting dataframe and creates a pie chart using these labels and sizes.



Word Clouds

Word clouds are a graphical representation of the most frequent words in a text, where the size of each word represents its frequency.

Python libraries such as word cloud and matplotlib were used to create these visualizations, which provide a quick and intuitive understanding of the main topics discussed in the text.

Results

NLP Vader Lexicon

This code groups the data by the company column and counts the number of articles for each company.
Then, it loops over the groups and creates a new dataframe for each group. For each group, it calculates the average sentiment score for both the title and summary columns using the `get_sentiment_scores` function.

```
Group: Adani
Title sentiment average: 0.03885966850828723
Summary sentiment average: 0.137321546961326
Group: Block
Title sentiment average: -0.18250787671232865
Summary sentiment average: -0.11688801369863006
Group: Blockchain
Title sentiment average: 0.07623582089552239
Summary sentiment average: 0.1319223880597015
Group: Ebix
Title sentiment average: 0.09495933333333328
Summary sentiment average: 0.25737599999999994
Group: Kandi
Title sentiment average: -0.007106172839506174
Summary sentiment average: 0.14979506172839502
Group: Loki
Title sentiment average: -0.03311818181818181
Summary sentiment average: 0.12353506493506491
Group: Lordstown
Title sentiment average: -0.04899999999999999
Summary sentiment average: 0.3453595744680851
Group: Nikola
Title sentiment average: 0.05930083333333334
Summary sentiment average: 0.2458983333333333
Group: Twitter
Title sentiment average: 0.03417256637168141
Summary sentiment average: 0.017971681415929214
```

Results

NLP MultinomialNB

Our code performs sentiment analysis on a dataframe `df` containing news articles by using the TextBlob library to calculate the polarity of the summary for each row.

The polarity values are then classified into positive, negative, or neutral using scikit-learn's CountVectorizer and a naive Bayes classifier. The user can then enter a new sentence to get the predicted sentiment using the same classifier and vectorizer.

Where 0 = Neutral, >0=Positive, <0 = Negative.

```
company     source
Adani       Hindenburg Research    0.014740
            Yahoo News            -0.081250
Block       Hindenburg Research    0.023624
            Yahoo News             0.056212
Blockchain  Hindenburg Research    0.028571
            Yahoo News             0.089069
Ebix        Hindenburg Research    0.014941
            Yahoo News             0.206528
Kandi       Hindenburg Research    0.032412
            Yahoo News             0.168561
Loki        Hindenburg Research    0.028571
            Yahoo News             0.071806
Lordstown   Hindenburg Research    0.041178
            Yahoo News             0.035000
Nikola      Hindenburg Research    0.060625
            Yahoo News             0.046875
Twitter     Hindenburg Research    0.082716
            Yahoo News             0.022386
Name: polarity, dtype: float64
```

Further Improvements

To improve the sentiment analysis model, there are a few approaches you can take:

- Increase the size and quality of the training data: A larger and more diverse dataset for training can improve the accuracy and robustness of the model.
- Fine-tune the model: Fine-tuning involves training the model on a smaller, domain-specific dataset to improve its performance on a specific task.
- Incorporate more advanced techniques: More advanced techniques, such as deep learning or reinforcement learning, can be used to improve the accuracy of the sentiment analysis model.
- Use a pre-trained model: There are pre-trained sentiment analysis models available that can be fine-tuned or used directly for specific tasks.
- As for alternatives to the current model, there are many other libraries and frameworks available for sentiment analysis, including VADER, Stanford CoreNLP, and spaCy. Each library or framework has its own strengths and weaknesses, so it's important to choose the one that best fits your specific use case and requirements.

Future Developments

- Employ web scraping tools to extract relevant data from financial news websites, and social media platforms
- Utilize cutting-edge NLP techniques such as recurrent neural networks, transformers and hyperparameter tuning
- Analyze and process textual data from various sources to derive sentiment scores
- Employ machine learning and deep learning techniques to classify and score sentiment data
- Employ advanced visualization techniques, such as heatmaps, network graphs, and sentiment score plots
- Leverage cloud-based solutions and distributed computing technologies for scalability
- Containerization technologies like Docker to package and deploy the project's components
- Ensure the project remains efficient and reliable as data volume and complexity increase
- Adapt to changes in market conditions, data sources, and investment strategies
- Incorporate feedback and new insights to continuously enhance the project's effectiveness

STEVENS
INSTITUTE OF TECHNOLOGY
1870

THANK YOU

Stevens Institute of Technology
1 Castle Point Terrace, Hoboken, NJ 07030