

Python for Data Science

Sentiment Analysis on the Belt & Road Initiative (BRI)



PRESENTED BY

22070126086 RHEA CHAINANI

22070126087 RICHA GUPTA

22070126098 SAMARTH PATEL

Identifying Problem

Belt and Road Initiative (BRI) is an infrastructural project launched by China to connect East Asia, Europe and Africa through physical infrastructure and aims at enhancing regional connectivity.

In this project, we aim to apply sentiment analysis to a diverse range of texts, including news articles and social media comments, to gain a comprehensive understanding of public sentiment towards the BRI.

Data Collection

We scraped YouTube comments on BRI related videos and combined that with the already existing dataset from Zenodo which contained headlines of Chinese newspaper starting from 2014 up till 2023. The dataset has 18191 entries of text, its recording year, source and corresponding sentiment labels and scores.

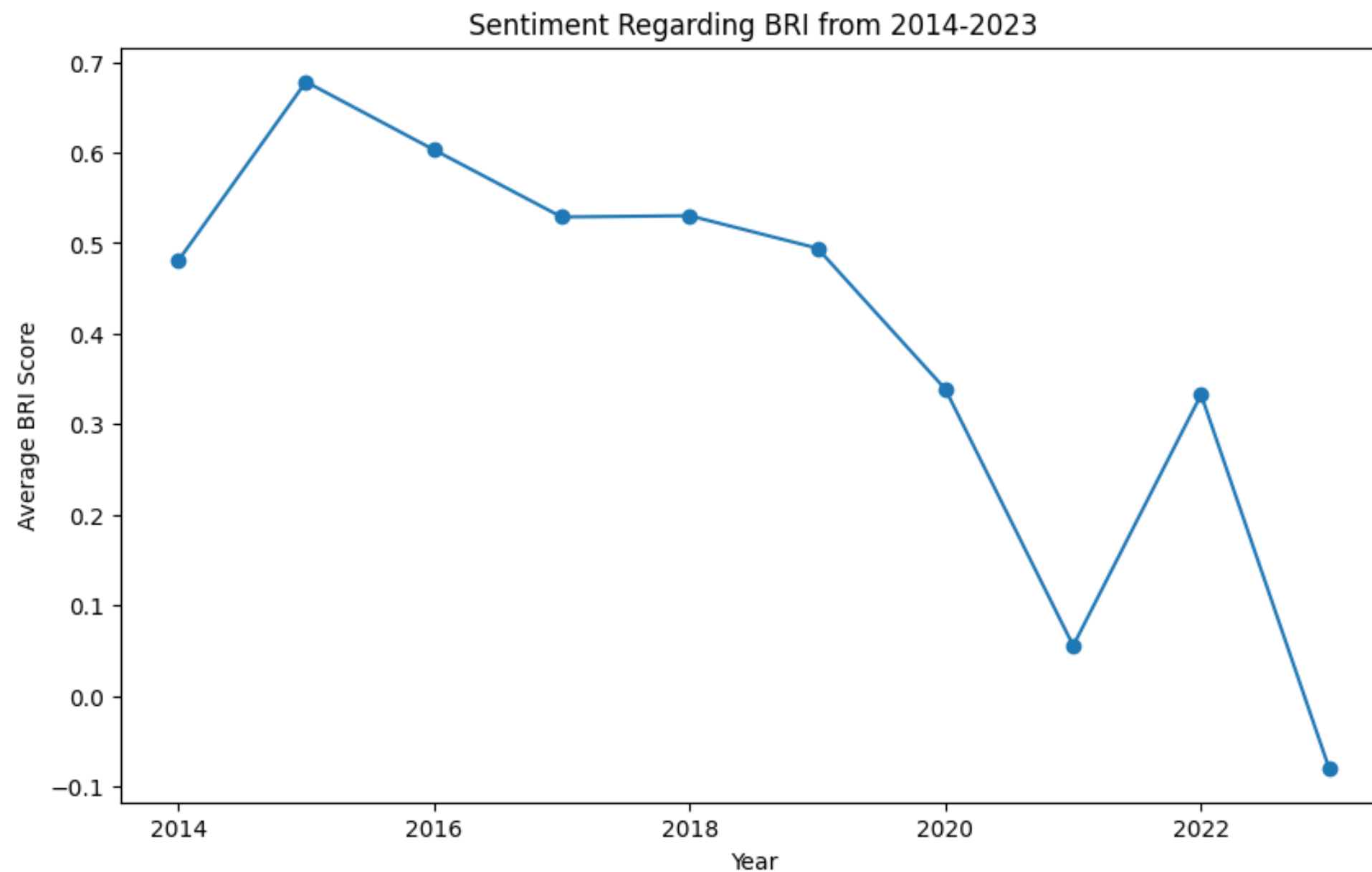
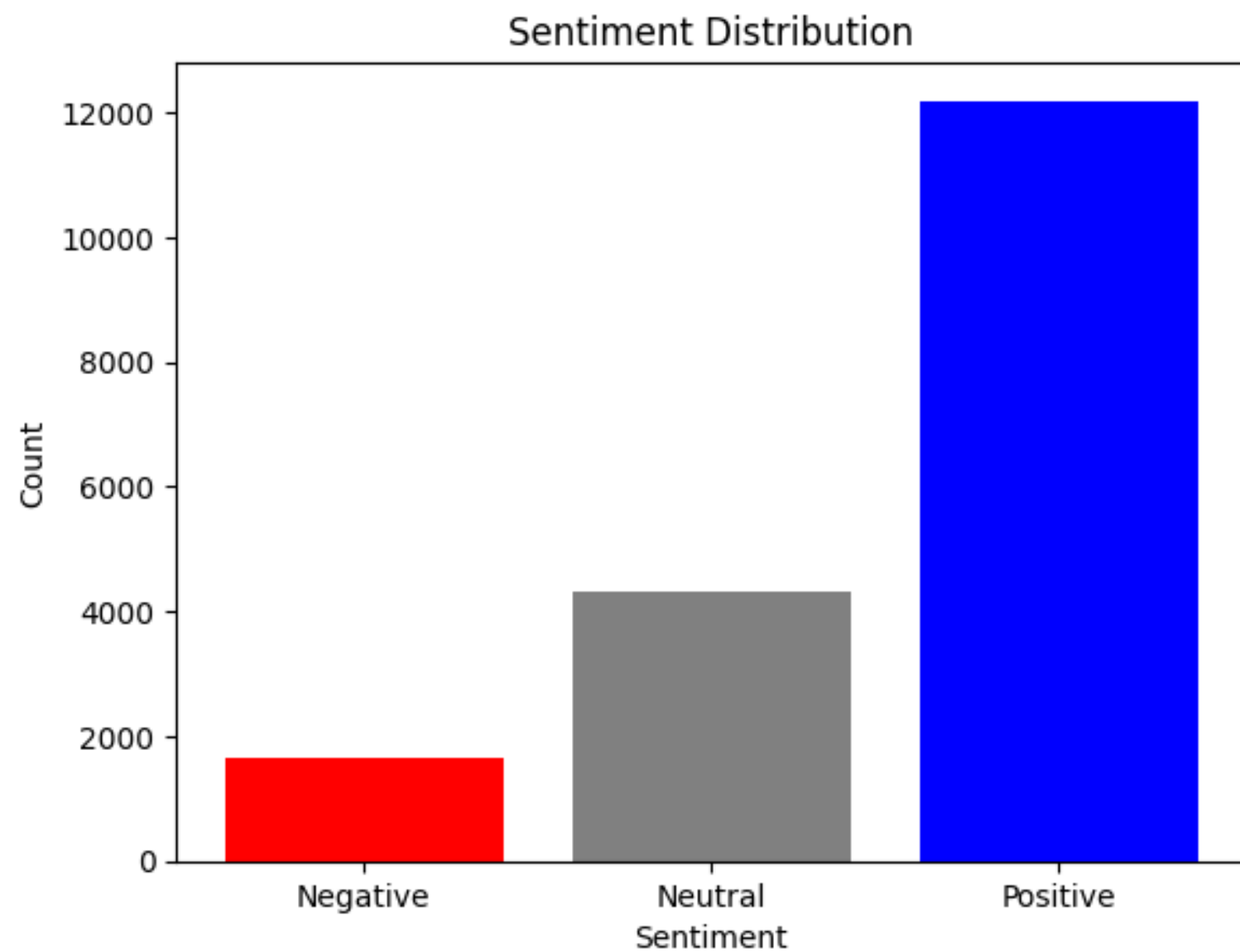
index	text	date	source	month	year	BRIscore	BRIsentiment	tokenized_text
0	yuan clearing service officially launched in argentina advancing internationalization	2022-11-09 23:34:00	globaltimes	11.0	2022.0	0.7351	Positive	['yuan', 'clearing', 'service', 'officially', 'launched', 'argentina', 'advancing', 'internationalization']
1	simulating speed	2022-11-08 22:56:00	globaltimes	11.0	2022.0	0.4215	Positive	['simulating', 'speed']
2	china and latin america have more and more common languages despite political changes in those countries	2022-11-09 21:51:00	globaltimes	11.0	2022.0	0.5994	Positive	['china', 'latin', 'america', 'common', 'languages', 'despite', 'political', 'changes', 'countries']
3	tackling climate change china and asean are important partners	2022-11-07 16:12:00	globaltimes	11.0	2022.0	0.4404	Positive	['tackling', 'climate', 'change', 'china', 'asean', 'important', 'partners']
4	chinalaos railway a vivid example of highquality joint bri construction between china and asean countries fm	2022-11-07 19:40:00	globaltimes	11.0	2022.0	0.9299	Positive	['chinalaos', 'railway', 'vivid', 'example', 'highquality', 'joint', 'bri', 'construction', 'china', 'asean', 'countries', 'fm']
5	over 90 of global businesses expect more trade with china under rcep survey	2022-11-07 21:48:00	globaltimes	11.0	2022.0	0.9355	Positive	['90', 'global', 'businesses', 'expect', 'trade', 'with', 'china', 'rcep', 'survey']

Data Processing

Various preprocessing steps were applied on the dataset:

1. Conversion to lower case
2. Removing punctuations
3. Handling emojis (for YouTube comments)
4. Removing non ascii characters (for Chinese characters in news headlines)
5. Stop word removal
6. Aspect based sentiment assigning
7. Lemmatization
8. Tokenization

Data Analysis



Data Modeling

We used Logistic Regression and got an accuracy of 71.9%. While using the data for modeling, we removed some of the positive sentiment texts in order to balance the data and get an unbiased and robust model.

```
BRIsentiment
Positive    12190
Neutral     4332
Negative    1669
Name: count, dtype: int64
```



```
BRIsentiment
Positive     7336
Neutral     4332
Negative    1669
Name: count, dtype: int64
```

Model	Accuracy
Logistic Regression	0.719640179910045
KNearestNeighbor	0.712143928035982
Naive Bayes	0.7068965517241379

Precision:

```
Macro Precision: 0.6825499563593512
Micro Precision: 0.719640179910045
Weighted Precision: 0.7082647227684463
```

f1-score:

```
Macro f1-score: 0.583525596272681
Micro f1-score: 0.719640179910045
Weighted f1-score: 0.6936435403196799
```

True Positive Rate:

```
True positive rate (sensitivity) for Negative: 0.19421487603305784
True positive rate (sensitivity) for Neutral: 0.5903975993998499
True positive rate (sensitivity) for Positive: 0.9148741418764302
```

False Positive Rate:

```
False positive rate (sensitivity) for Negative: 0.017055144968732235
False positive rate (sensitivity) for Neutral: 0.12289246908954665
False positive rate (sensitivity) for Positive: 0.4039625756741882
```

Sensitivity Analysis:

Text from Sample:

```
china is taking over everything when are we going to stop them
Sentiment for the Smaple text is: Negative
```

Modified Text:

```
china is taking over everything, we must encourage them
Sentiment for the Modified text is: Positive
```

Recall:

```
Macro Recall: 0.5664955391031127
Micro Recall: 0.719640179910045
Weighted Recall: 0.719640179910045
```

Jaccard's Index:

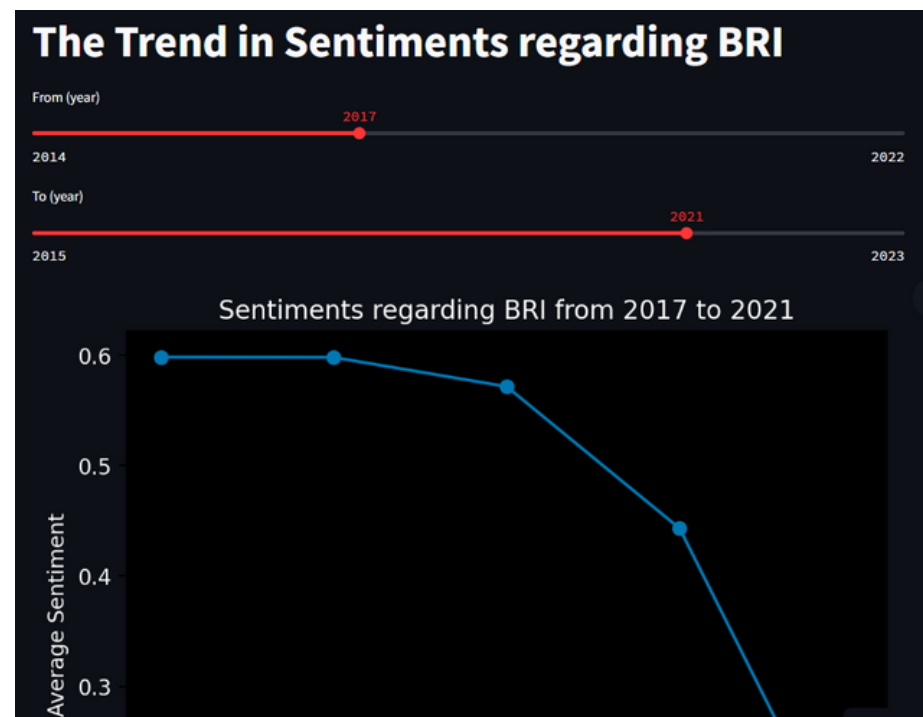
```
Macro Jaccard's Index: 0.44380954820579355
Micro Jaccard's Index: 0.5620608899297423
Weighted Jaccard's Index: 0.5526141576187793
```

Confusion Matrix:

```
[[ 94  163  227]
 [ 39  787  507]
 [ 21  165 1999]]
```


Model Deployment

Our model was globally deployed using Streamlit and Github.



What is the Belt & Road Initiative ?

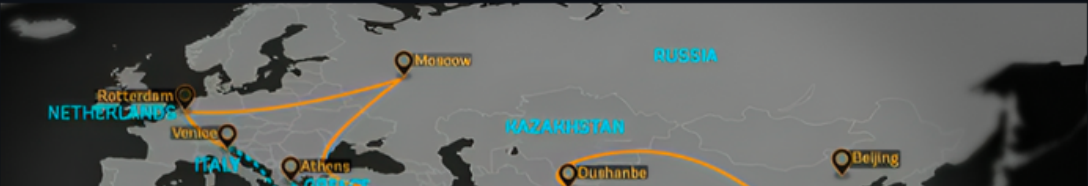
Select a section to check out

- ☒ About BRI
- ☐ Predict Sentiment of a Text
- ☐ View the Trends over the years

China's Belt and Road Initiative (BRI), sometimes referred to as the New Silk Road, is one of the most ambitious infrastructure projects ever conceived. Launched in 2013 by President Xi Jinping, the vast collection of development and investment initiatives was originally devised to link East Asia and Europe through physical infrastructure. In the decade since, the project has expanded to Africa, Oceania, and Latin America, significantly broadening China's economic and political influence.

Some analysts see the project as an unsettling extension of China's rising power, and as the costs of many of the projects have skyrocketed, opposition has grown in some countries. Meanwhile, the United States shares the concern of some in Asia that the BRI could be a Trojan horse for China-led regional development and military expansion. President Joe Biden has maintained his predecessors' skeptical stance towards Beijing's actions, but Washington has struggled to offer participating governments a more appealing economic vision.

China's overall ambition for the BRI is staggering. To date, 147 countries—accounting for two-thirds of the world's population and 40 percent of global GDP—have signed on to projects or indicated an interest in doing so.



[Link](#)

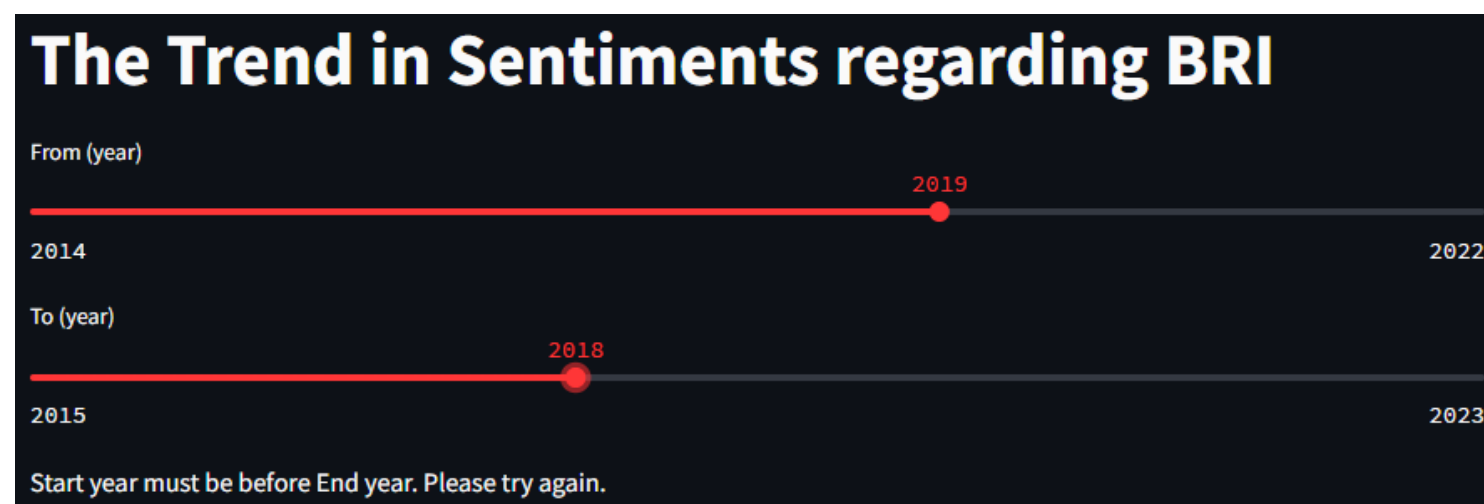
Predict Sentiment of a Text

Enter text

BRI is a success

Predict

😊 The sentiment of the text is : Positive



Predict Sentiment of a Text

Enter text

BRI is a failure

Predict

😞 The sentiment of the text is : Negative

Predict Sentiment of a Text

Enter text

BRI is nothing

Predict

😐 The sentiment of the text is : Neutral

Thank
You