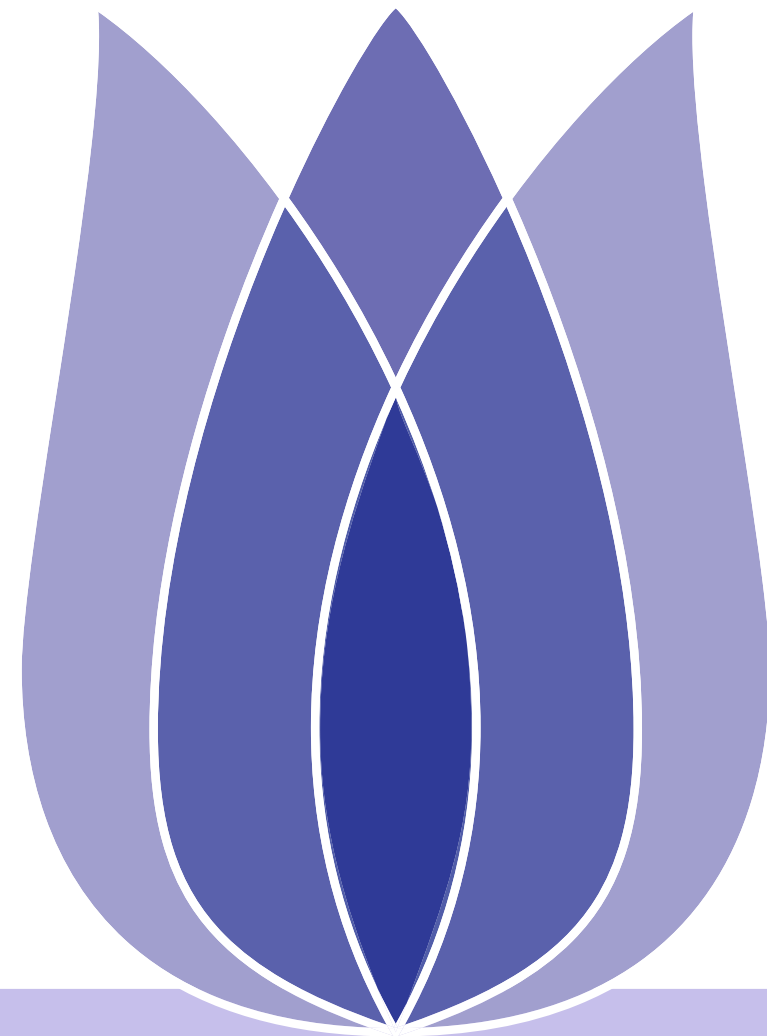


FLIP01 Final Assessment

Cong Ma

QingDao Technological University

2021-01-16





Overview

- [Problem Definition](#)
- [Data Visualization](#)
- [Data Process](#)
- [Build The Model](#)
- [Conclusion](#)

Problem Definition

Problem Description

Data Visualization

Data Visualization

Data Process

Data Process

Build The Model

Model:CRF+LSTM

Model:roBERTa

Conclusion

Contact Information



Problem Definition

Problem Description

Data Visualization

Data Process

Build The Model

Conclusion

Problem Definition



Problem Description

Problem Definition
Problem Description
Data Visualization
Data Process
Build The Model
Conclusion

Defn

With all of the tweets circulating every second it is hard to tell whether the sentiment behind a specific tweet will impact a company, or a person’s, brand for being viral (positive), or devastate profit because it strikes a negative tone.

- What’s the **Sentiment** of this tweet.
- What’s the part of the tweet (**word or phrase**) that reflects the sentiment.

ID	text	selected_text	sentiment
<i>cb774db0d1</i>	Uh oh, I am sunburned	I am sunburned	negative
<i>549e992a42</i>	We saw that the baddie’s the best	best	positive
<i>f84b89a828</i>	Sounds like me	Sounds like me	neutral



[Problem Definition](#)

[Data Visualization](#)

[Data Visualization](#)

[Data Process](#)

[Build The Model](#)

[Conclusion](#)

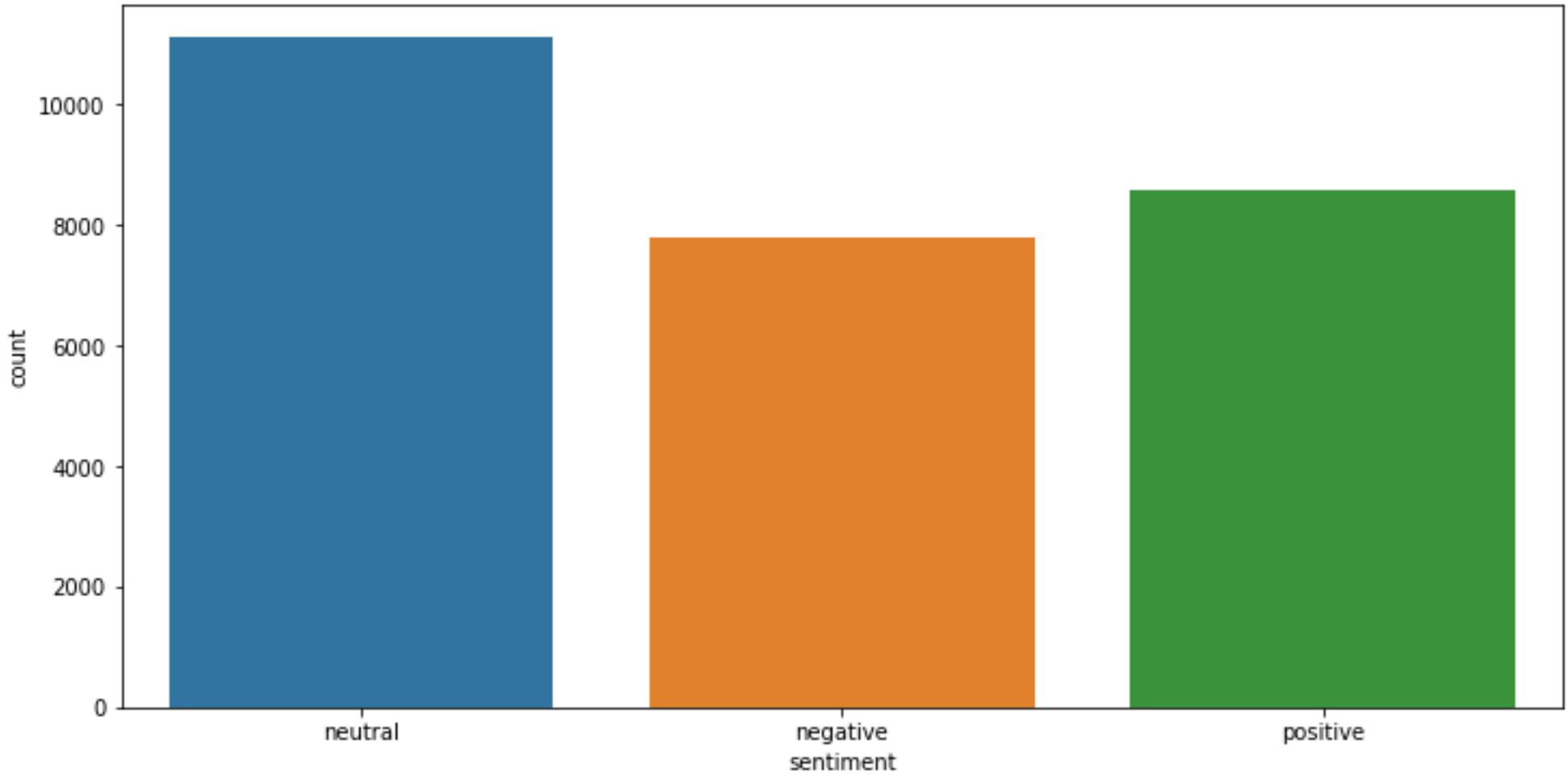
Data Visualization



Data Visualization

- Problem Definition
- Data Visualization
- Data Visualization
- Data Process
- Build The Model
- Conclusion

- First, check the data. The training set contains 27482 data.
 - ◆ Take a look at the proportion of different types of text in the training set
 - ◆ It can be seen that the number of three kinds of data is relatively average. In addition, there are more neutral texts.

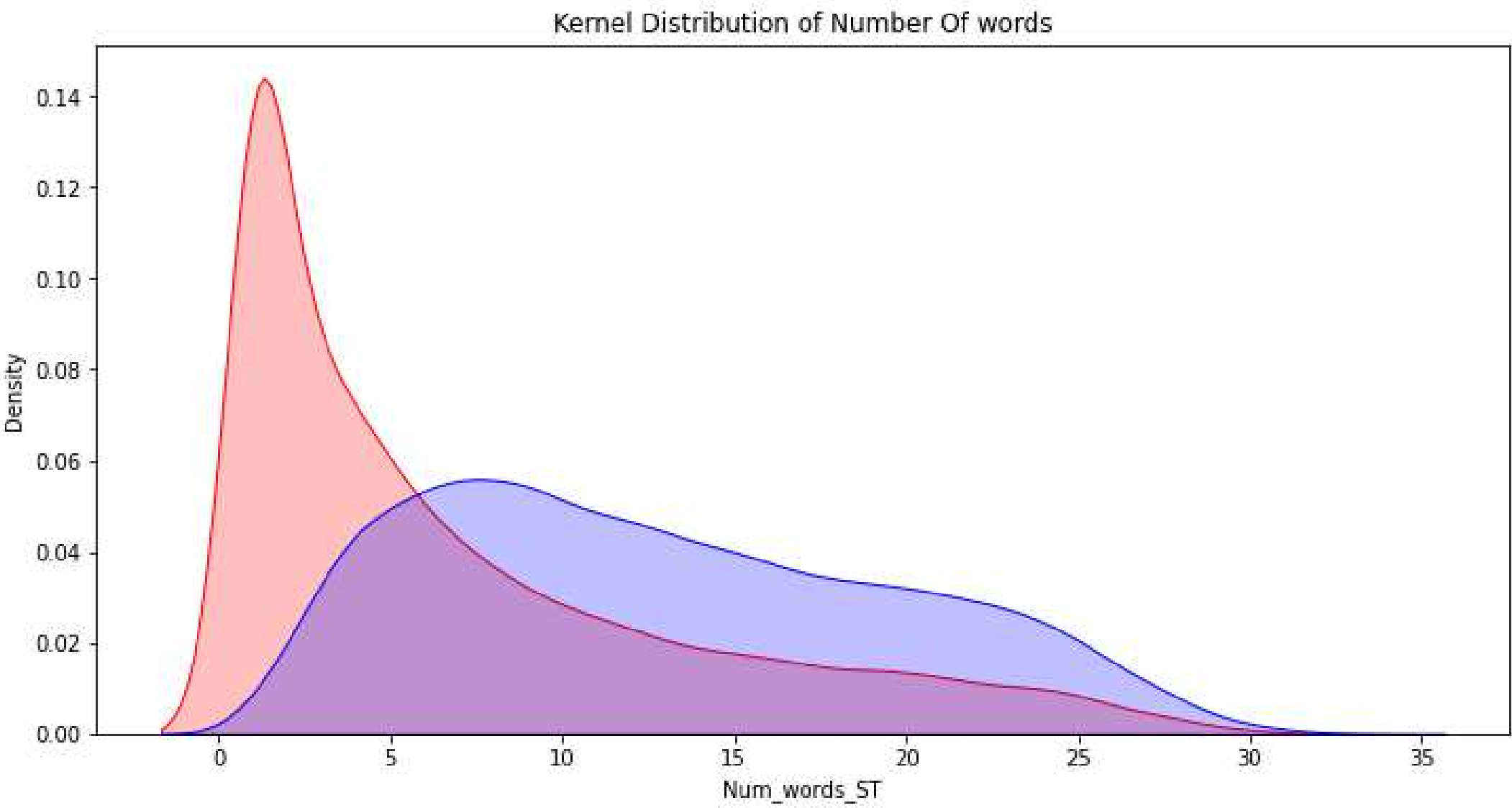




Data Visualization

- [Problem Definition](#)
- [Data Visualization](#)
- [Data Visualization](#)
- [Data Process](#)
- [Build The Model](#)
- [Conclusion](#)

- Count the distribution interval of the length of the given text and the selected text.

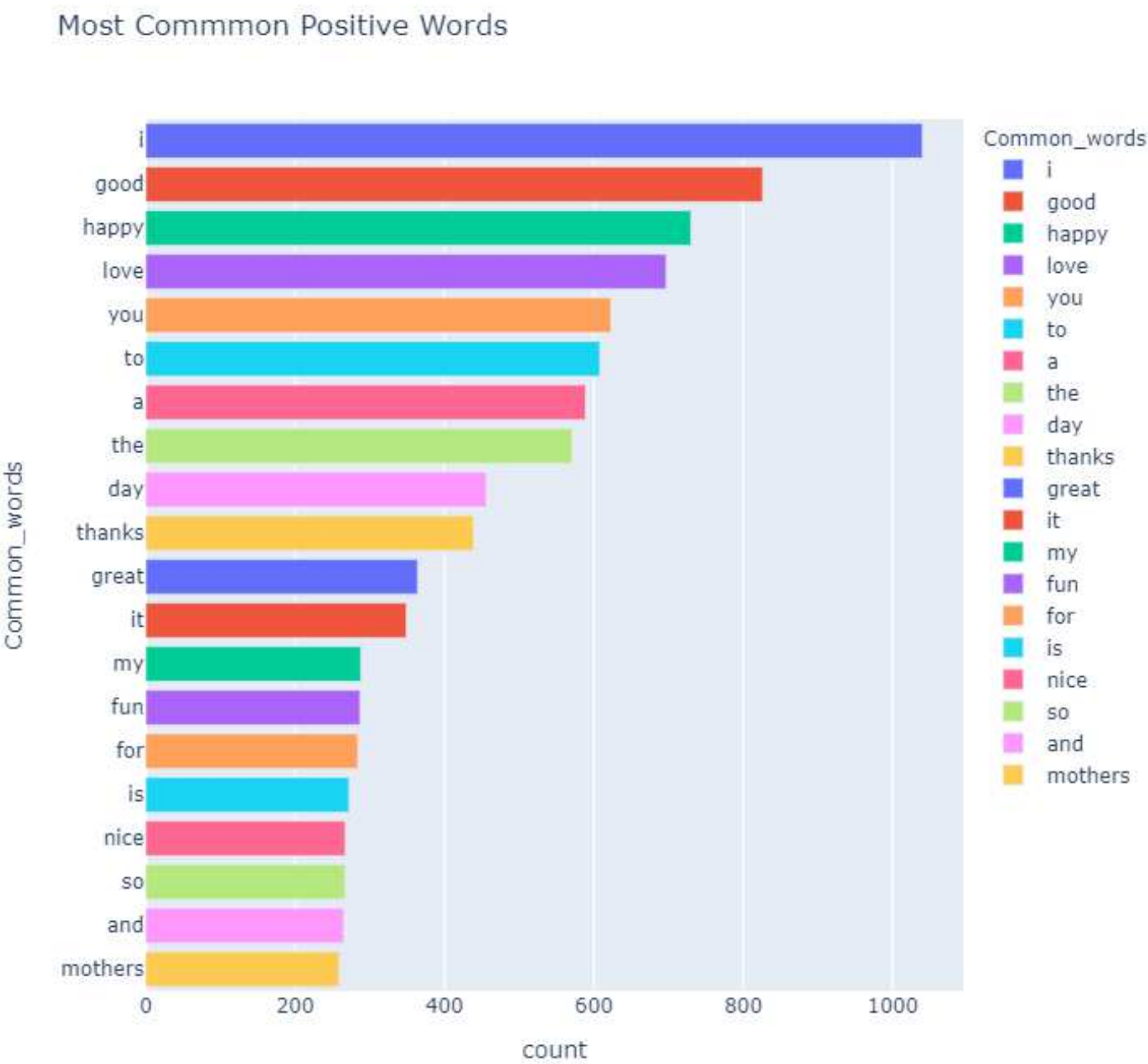




Data Visualization

- Problem Definition
- Data Visualization
- Data Visualization**
- Data Process
- Build The Model
- Conclusion

- Statistics of positive emotions were selected in the text of the highest frequency of the first few words.





- Problem Definition
- Data Visualization
- Data Visualization
- Data Process
- Build The Model
- Conclusion

- The statistical results will be generated word cloud to more intuitive look at the frequency of words.

WordCloud of Postive Tweets

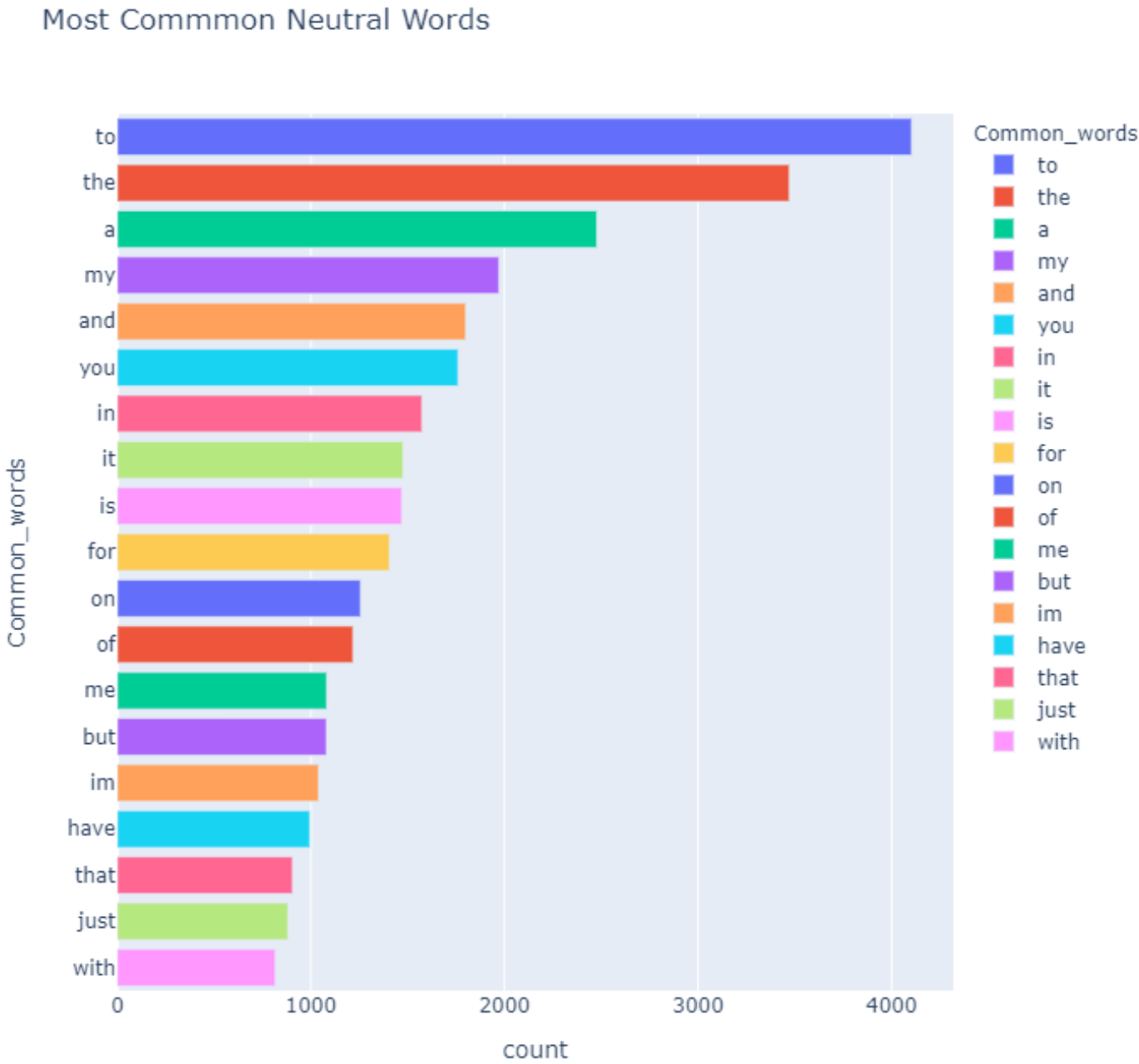




Data Visualization

- Problem Definition
- Data Visualization
- Data Visualization**
- Data Process
- Build The Model
- Conclusion

- Statistics of neutral emotions were selected in the text of the highest frequency of the first few words.





Problem Definition

Data Visualization

Data Visualization

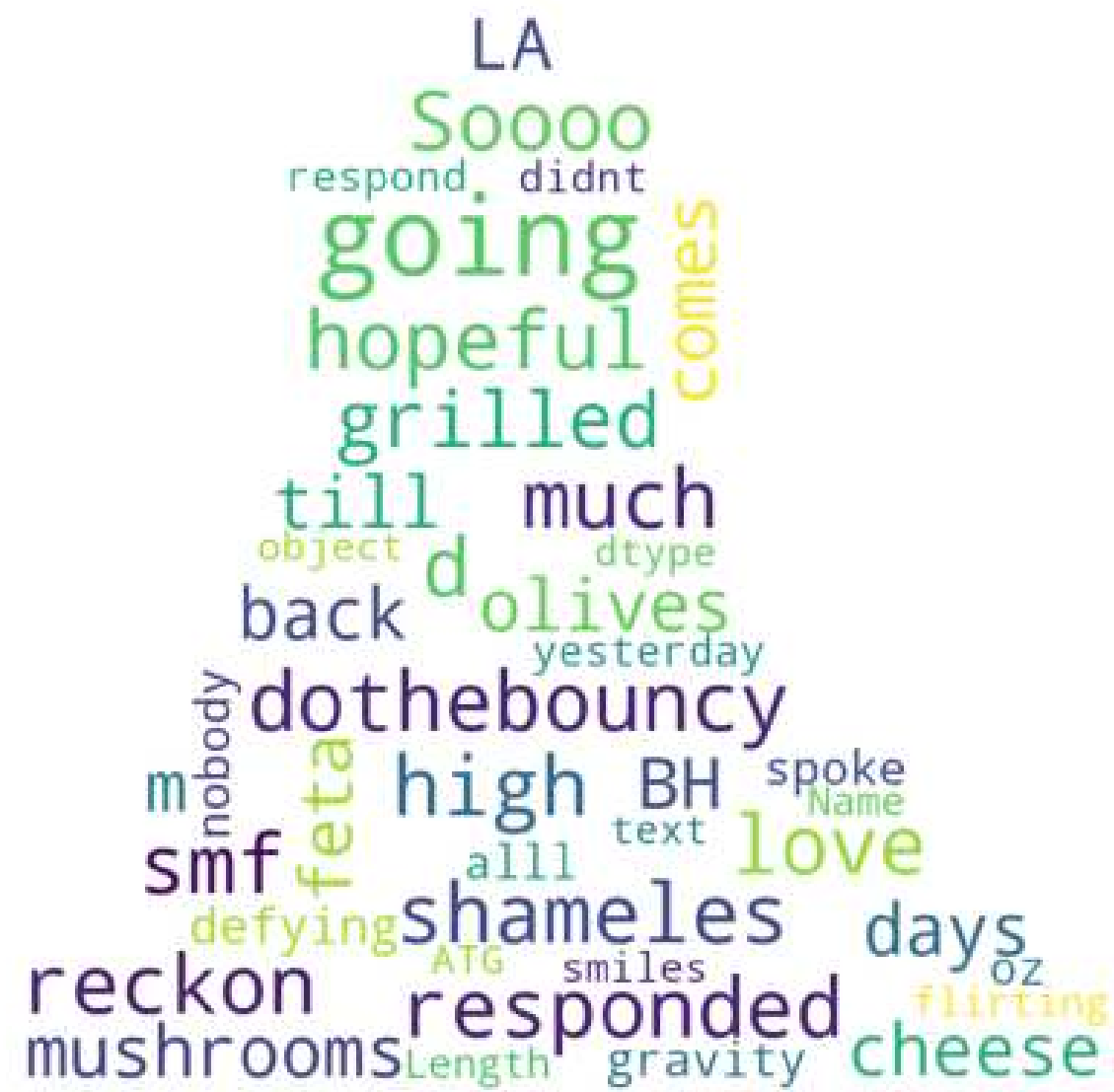
Data Process

Build The Model

Conclusion

- The statistical results will be generated word cloud to more intuitive look at the frequency of words.

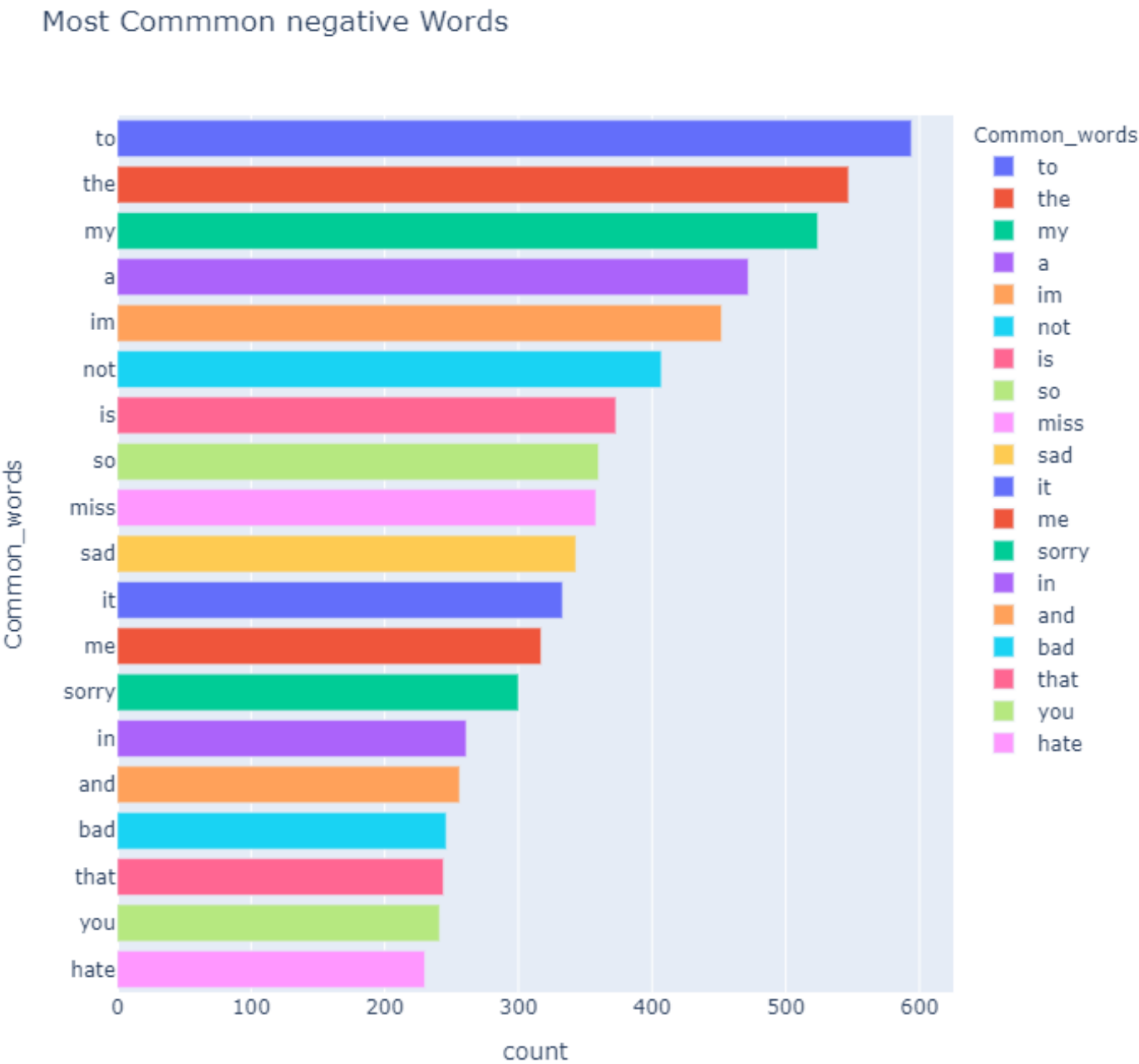
WordCloud of Neutral Tweets





- Problem Definition
- Data Visualization
- Data Visualization**
- Data Process
- Build The Model
- Conclusion

- Statistics of negative emotions were selected in the text of the highest frequency of the first few words.





- Problem Definition
- Data Visualization
- Data Visualization**
- Data Process
- Build The Model
- Conclusion

- The statistical results will be generated word cloud to more intuitive look at the frequency of words.

WordCloud of negative Tweets





Data Visualization

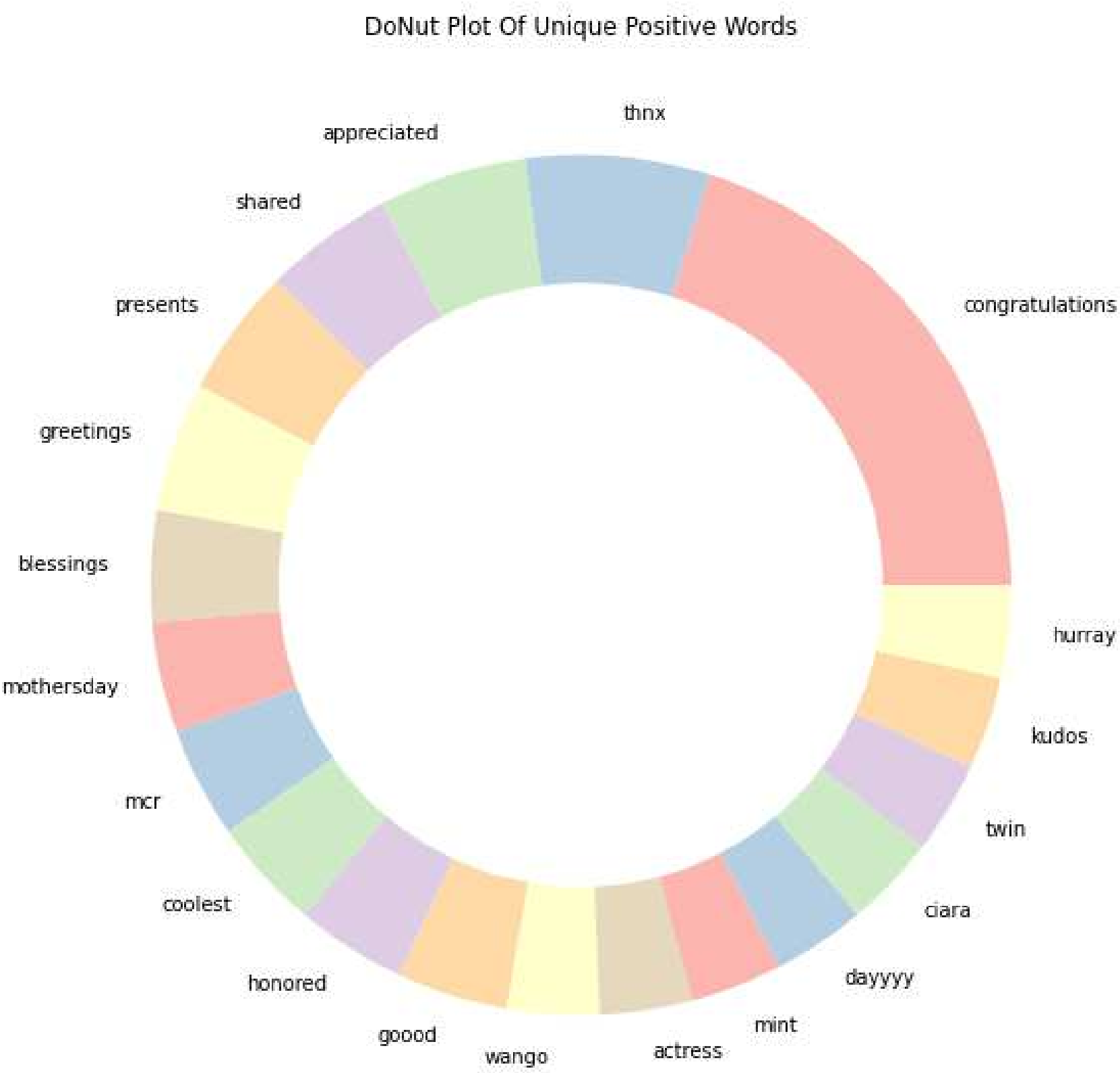
- [Problem Definition](#)
- [Data Visualization](#)
- [Data Visualization](#)
- [Data Process](#)
- [Build The Model](#)
- [Conclusion](#)

- It can be seen that our previous statistical text contains some words without emotional tendency.
- After we delete these words, we count the frequency of each word.



Data Visualization

- Problem Definition
- Data Visualization
- Data Visualization**
- Data Process
- Build The Model
- Conclusion

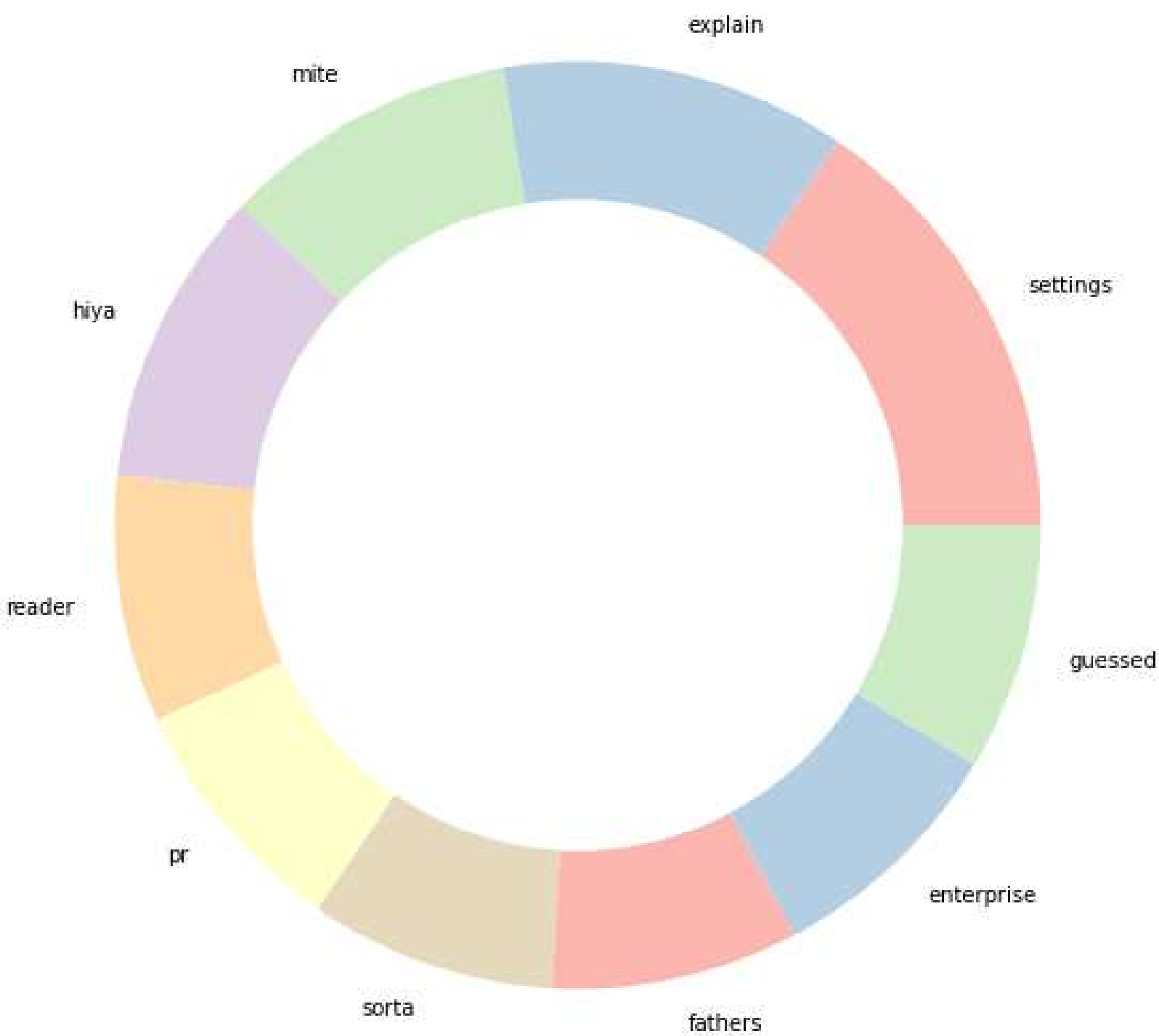




Data Visualization

- [Problem Definition](#)
- [Data Visualization](#)
- [Data Visualization](#)
- [Data Process](#)
- [Build The Model](#)
- [Conclusion](#)

DoNut Plot Of Unique Neutral Words

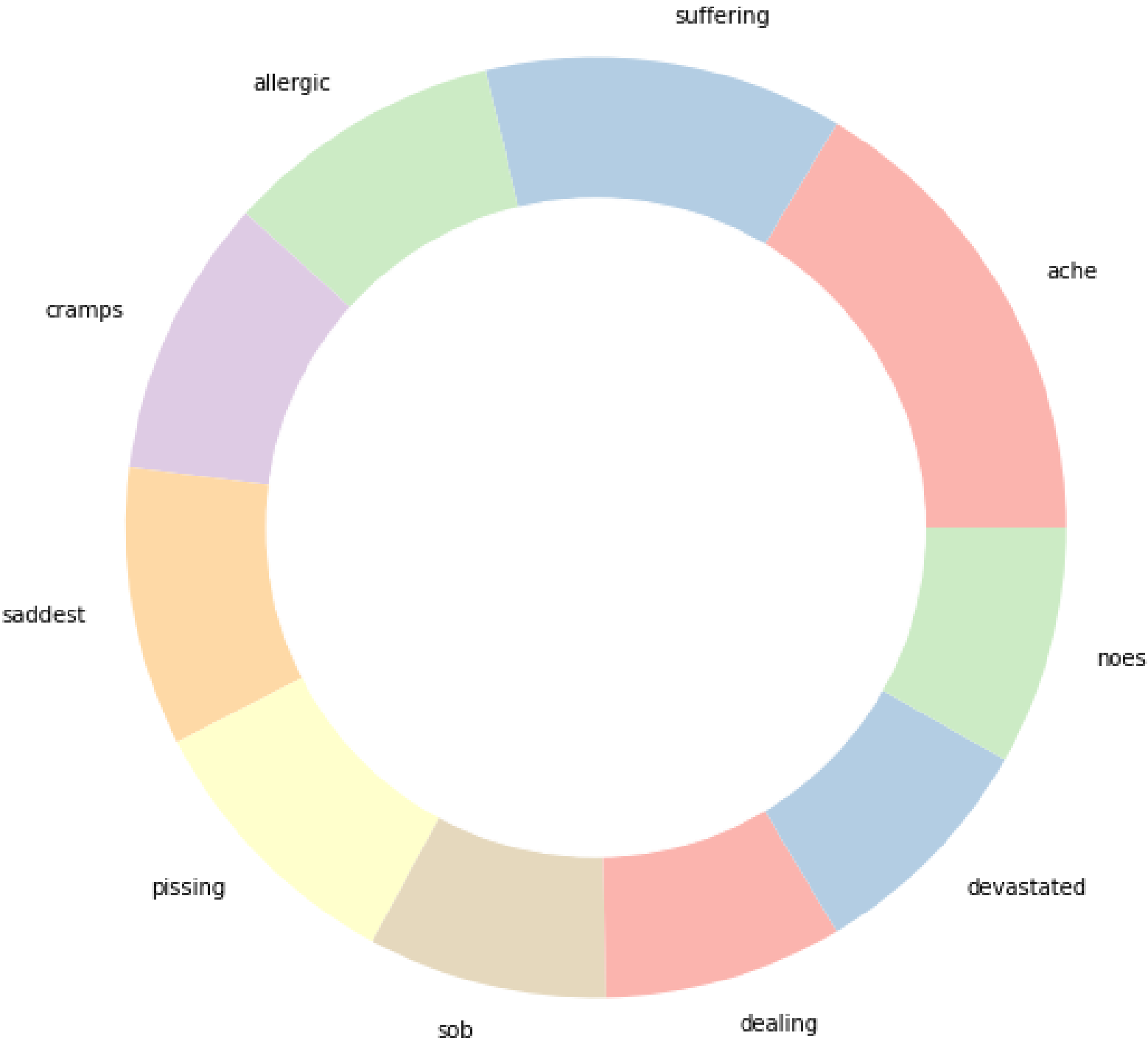




Data Visualization

- [Problem Definition](#)
- [Data Visualization](#)
- [Data Visualization](#)
- [Data Process](#)
- [Build The Model](#)
- [Conclusion](#)

DoNut Plot Of Unique Negative Words





[Problem Definition](#)

[Data Visualization](#)

[Data Process](#)

[Data Process](#)

[Build The Model](#)

[Conclusion](#)

Data Process



Data Process

- [Problem Definition](#)
- [Data Visualization](#)
- [Data Process](#)
- [Data Process](#)
- [Build The Model](#)
- [Conclusion](#)

- Observe the given training set, and the extracted words are part of the original sentence.
- The data processing part will only delete the blank data in the given training set.



[Problem Definition](#)

[Data Visualization](#)

[Data Process](#)

[Build The Model](#)

Model:CRF+LSTM

Model:roBERTa

[Conclusion](#)

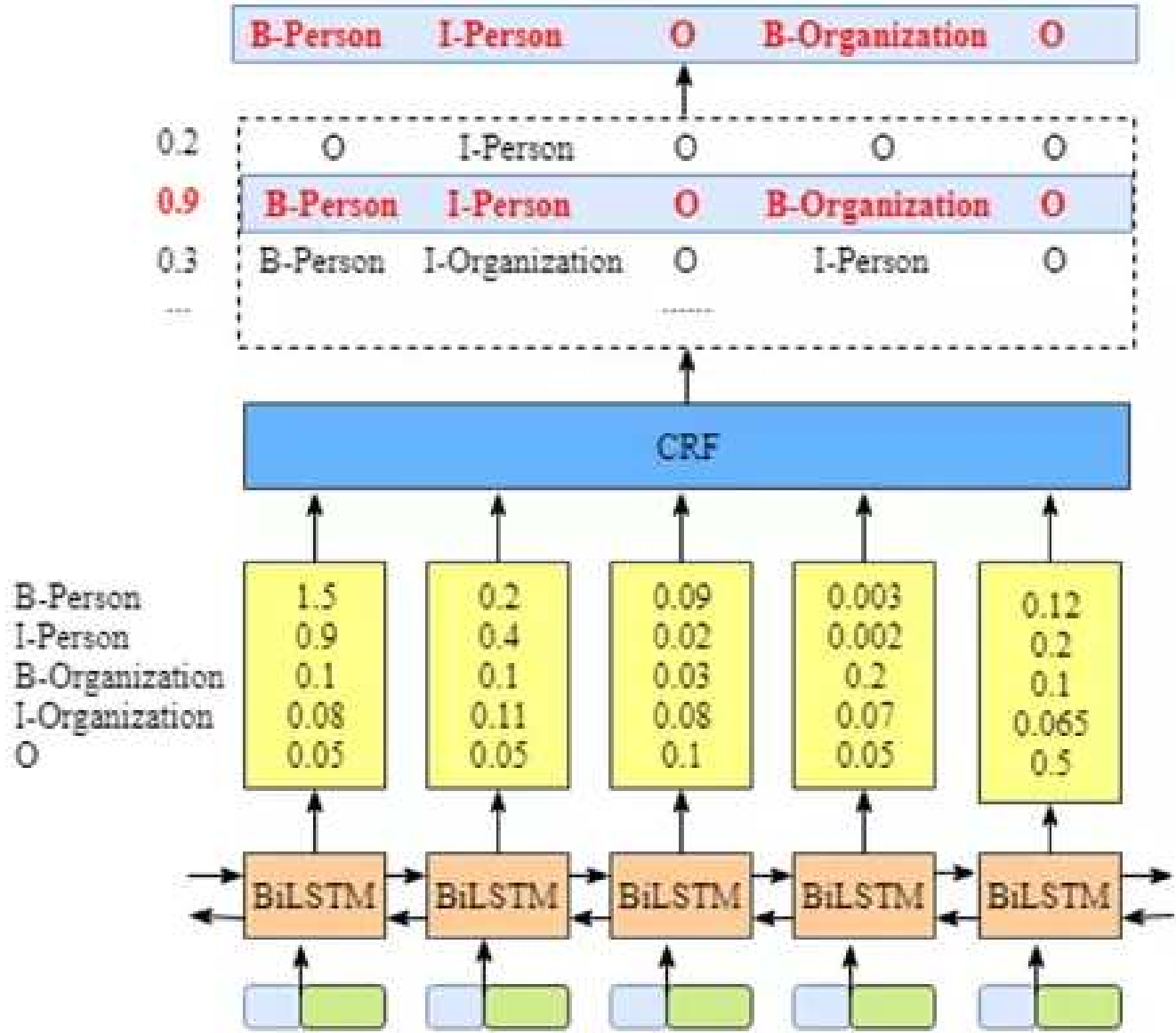
Build The Model



Model:CRF+LSTM

- Problem Definition
- Data Visualization
- Data Process
- Build The Model
- Model:CRF+LSTM**
- Model:roBERTa
- Conclusion

- First, the CRF+LSTM model is used.





Model:CRF+LSTM

Problem Definition
Data Visualization
Data Process
Build The Model
Model:CRF+LSTM
Model:roBERTa
Conclusion

- Based on our previous data visualization.We set the MAX_LEN = 48.
- The learning rate is 0.8.
- Activation function is “Softmax”.
- The loss function is “categorical_crossentropy”.
- Epochs = 10.





Model:CRF+LSTM

- [Problem Definition](#)
- [Data Visualization](#)
- [Data Process](#)
- [Build The Model](#)
- [Model:CRF+LSTM](#)**
- [Model:roBERTa](#)
- [Conclusion](#)

	textID	text	selected_text	sentiment
0	d93afa85cf	Car not happy, big big dent in boot! Hoping t...	Car not happy, big big dent in boot! Hoping th...	neutral
1	fab6b7d16c	im an avid fan of **** magazine and i love you...	avid fan of	positive
2	2e7082d1c8	MAYDAY?!	MAYDAY?!	neutral
3	684081e4e7	RATT ROCKED NASHVILLE TONITE..ONE THING SUCKED...	RATT ROCKED NASHVILLE TONITE..ONE THING SUCKED...	neutral
4	c77717b103	I love to! But I'm only available from 5pm. ...	I love to!	positive

- Finally, the accuracy of the trained model is 0.5393.



Model:roBERTa

[Problem Definition](#)

[Data Visualization](#)

[Data Process](#)

[Build The Model](#)

[Model:CRF+LSTM](#)

[Model:roBERTa](#)

[Conclusion](#)

In order to obtain higher accuracy, I choose the widely used model named roBERTa.

- Roberta: a robust method to optimize the pre training of Bert.
- Roberta is an improved algorithm of bert.
 - ◆ With bigger batchsize and more data, let the model train longer.
 - ◆ Removed the NSP (next sense prediction) task.
 - ◆ Mask mechanism for dynamically modifying training data.



TULIP

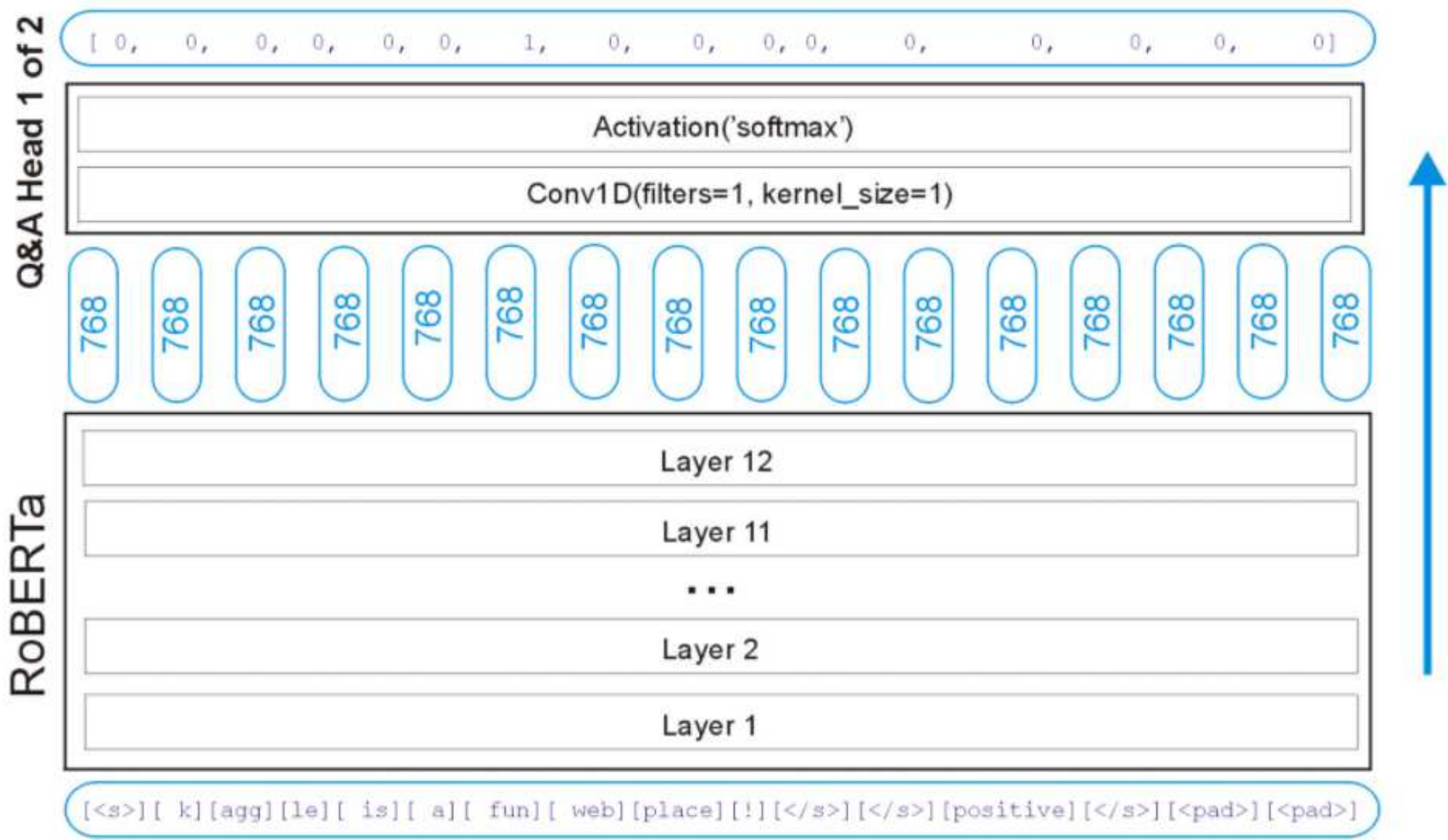
Team for Universal Learning and Intelligent Processing



Model:roBERTa

- Problem Definition
- Data Visualization
- Data Process
- Build The Model
- Model:CRF+LSTM
- Model:roBERTa
- Conclusion

- Activation function is “softmax”.
- The output is one dimension and the convolution kernel size is 1 * 1.





Model:roBERTa

- [Problem Definition](#)
- [Data Visualization](#)
- [Data Process](#)
- [Build The Model](#)
- [Model:CRF+LSTM](#)
- [Model:roBERTa](#)
- [Conclusion](#)

- The learning rate is 0.9.
- The optimizer is “Adam”.
- The loss function is “categorical_crossentropy”.
- Using k-fold cross validation, it is divided into five parts. Train three times.



Model:roBERTa

- Problem Definition
- Data Visualization
- Data Process
- Build The Model
 - Model:CRF+LSTM
 - Model:roBERTa
- Conclusion

	textID	text	selected_text	sentiment
0	eae9c20c8d	#followfriday thank you so much. I`m so be...	day thank you so mu	positive
1	404e86f215	(cont) when told him that I love beans on toa...	I love beans on toast. SO CUTE!	positive
2	81a83e8d9a	why am i up so early	why am i up so early	negative
3	c0d5b45663	Joined you on facebook!	Joined you on facebook!	neutral
4	c3c1abb017	trying to find some friends and not having any...	not having any luck	negative

- Finally, the accuracy of the trained model is 0.7072782233322157.



- [Problem Definition](#)
- [Data Visualization](#)
- [Data Process](#)
- [Build The Model](#)
- [Conclusion](#)**
- [Contact Information](#)

Conclusion



Contact Information

Made By Cong Ma
QingDao Technological University

