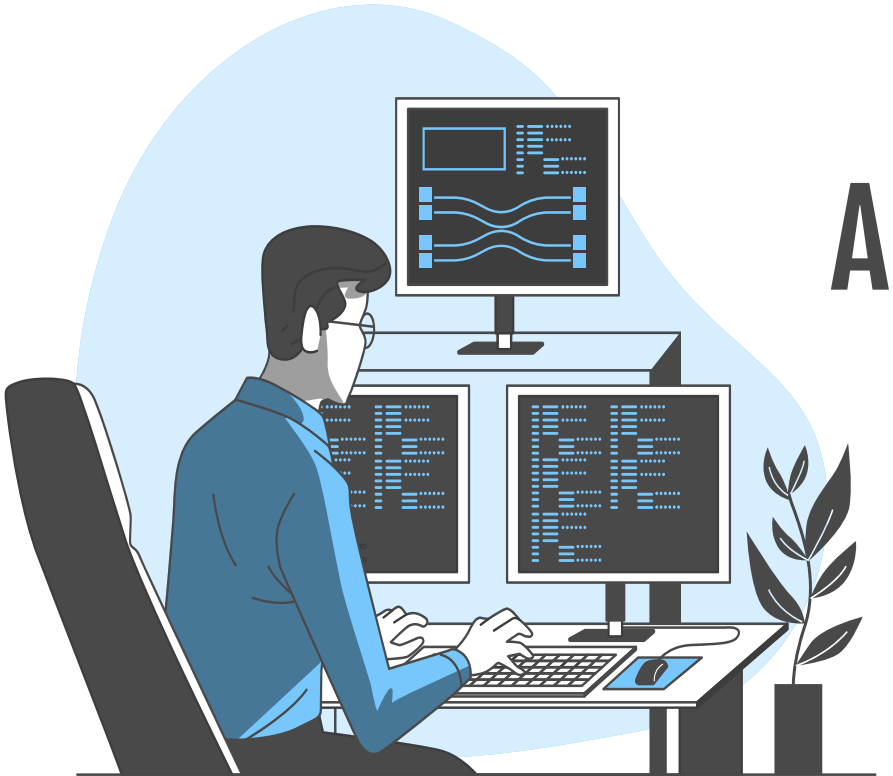
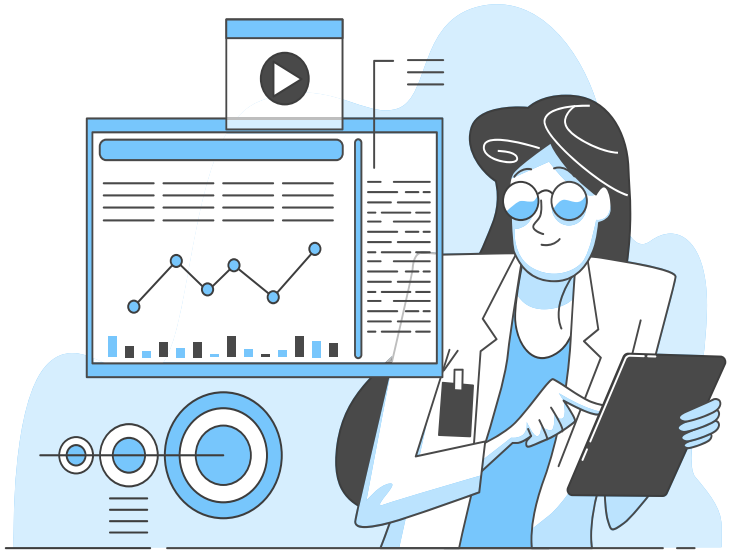


Sentiment Analysis of Reddit Comments

For Pre-Identified Targets

Table of Contents



01
Introduction

Introduction

- Social media sites like Reddit, Facebook, Twitter are sites where users can express their views on various topics .
- Large number of users often express their opinions on politicians/policy on such sites.
- Sentiment analysis can be used to gauge public opinion on policy announcements/ politician approval in faster traditional methods like panels, surveys or studies. Allowing for quicker implementation or response times.

Pipeline

01

Identification of Targets

- Identify Target entities
- Identify aliases and other references
- Unify all references under a single label

02

Break-up comments with multiple entities

- Sentiment labels/scores are normally a compound score given to an entire comment
- Breakdown Comments with multiple entities to clauses

03

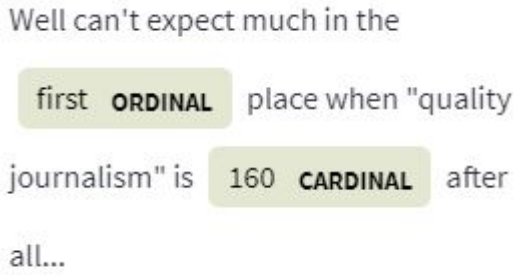
Assign a sentiment to the comment (or parts)

- Analyse comment and classifying the comment

02
The Model

01 - Identifying Target entities

- Existing NER models face some difficulty in recognising words that have alternate meanings.

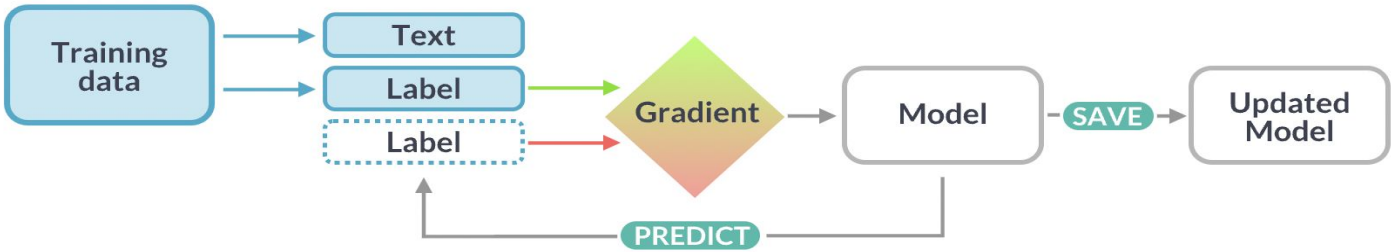


- Target entities are often referred to in a variety of ways:
  - Initials - LHL
  - First names - Amy, Jamus
  - Short forms - Jo Teo
  - Nicknames - Kee Chiu

01 - Model Training - Spacy

- Customised Named Entity Recognition ('NER') and Named Entity Linker ('NEL') Components
- NER identifies any named entities in a text
- NEL determines if the identified entities are relevant

01 - Model Training - Spacy



01 - Training Data

- Annotated data
- Utilised a rule based approach to label comments with the entity
- re-label/ remove any mislabeled items

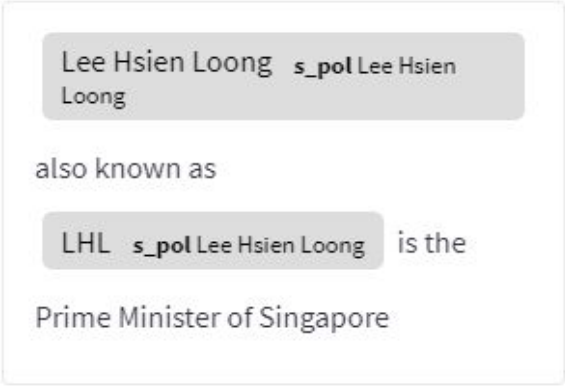
Text	Label
Singapore Politics is a simple game; several parties chase votes for a few weeks every 5 years, and at the PAP wins.	Start Char: 108 End Char: 111 Label: ORG Entity: People's Action Party
NOC Co-Founder Sylvia Chan was diagnosed with severe depression, OCD & rage disorder when she was In JC (unapologetic and sue all your ass syndrome too)	Start Char: 15 End Char: 21 Label: S_Pol Entity: Slyvia Lim

01 - Performance

	Ents_Percision	Ents_Recall	Ents_Weighted F1
NER	0.79	0.98	0.88
NEL	-	-	0.98

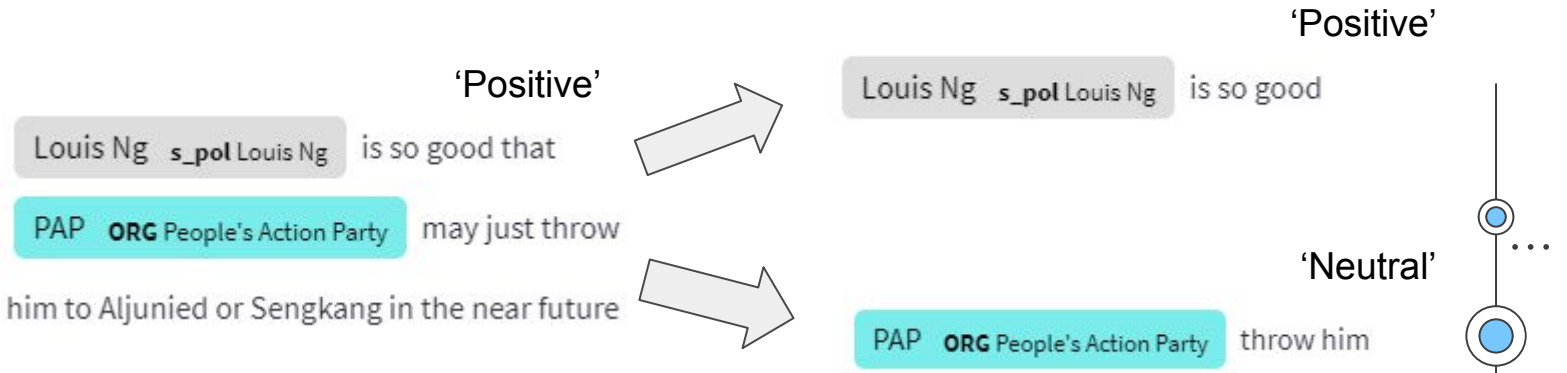
01 - Identifying Target entities

- With the custom components we can:
  - Localise context
  - Group difference references of an individual under a single entity
  - Filter out entities that we are uninterested in.



02 – Multiple Entities referenced

- Comments may include multiple entities
- Difficult to assign a score to to all entities in an equitable manner

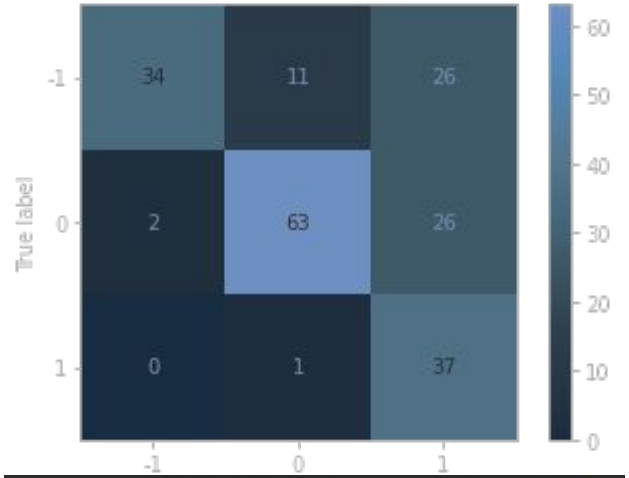


03- Determining Comment Sentiment

Model	Weighted - F1
Textblob*	0.3
Vader	0.68
Pre-trained Classification Models (Roberta)	0.87

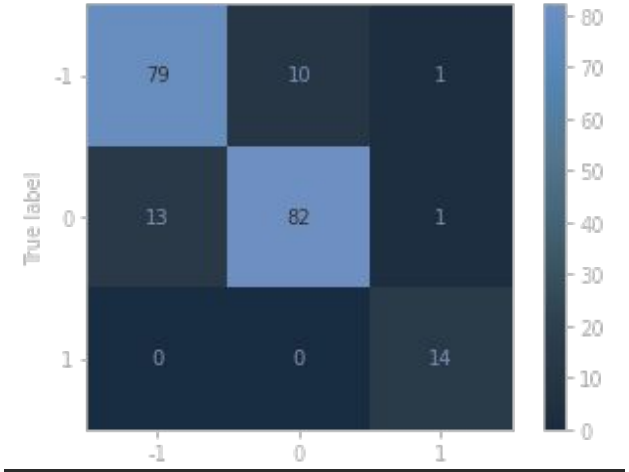
03- Determining Comment Sentiment

Vader



	precision	recall	f1-score	support
negative	0.94	0.48	0.64	71
neutral	0.84	0.69	0.76	91
positive	0.42	0.97	0.58	38
accuracy			0.67	200
macro avg	0.73	0.71	0.66	200
weighted avg	0.80	0.67	0.68	200

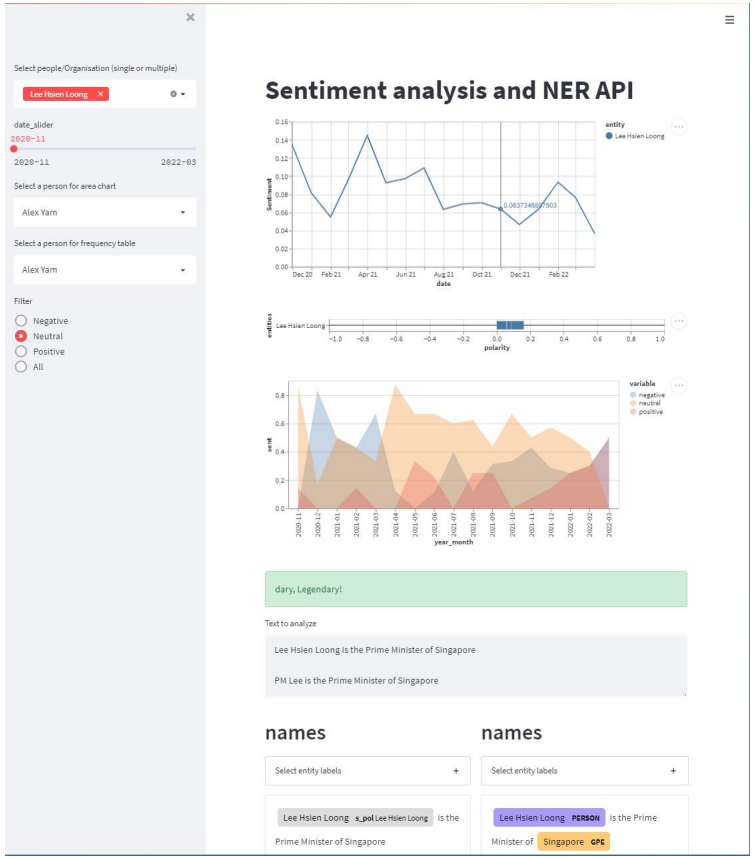
cardiffnlp/twitter-roberta-base-sentiment



	precision	recall	f1-score	support
negative	0.86	0.88	0.87	90
neutral	0.89	0.85	0.87	96
positive	0.88	1.00	0.93	14
accuracy			0.88	200
macro avg	0.88	0.91	0.89	200
weighted avg	0.88	0.88	0.87	200

03
Dashboard

Dashboard


Select people/Organisation (single or multiple)
Lee Hsien Loong
date_slider
2020-11
2020-11
2022-03
Select a person for area chart
Alex Yam
Select a person for frequency table
Alex Yam
Filter
Negative
Neutral
Positive
All
Sentiment analysis and NER API
entity
Lee Hsien Loong
Sentiment
date
entities
Lee Hsien Loong
polarity
variable
negative
neutral
positive
sent
year_month
dary, Legendary!
Text to analyze
Lee Hsien Loong is the Prime Minister of Singapore
PM Lee is the Prime Minister of Singapore
names
Select entity labels
Lee Hsien Loong s_pol Lee Hsien Loong is the Prime Minister of Singapore
names
Select entity labels
Lee Hsien Loong PERSON is the Prime Minister of Singapore GPE

Conclusion

Conclusion

Highlights

- ~88% F1 Score
- ~98% Recall

Limitations

- English only
- Cannot identify what aspect people are reacting to.

Future Improvements

- Customised word embeddings for words outside of vocab / alternate usage.
- Additional component to classify topics

Thanks!

Do you have any questions?