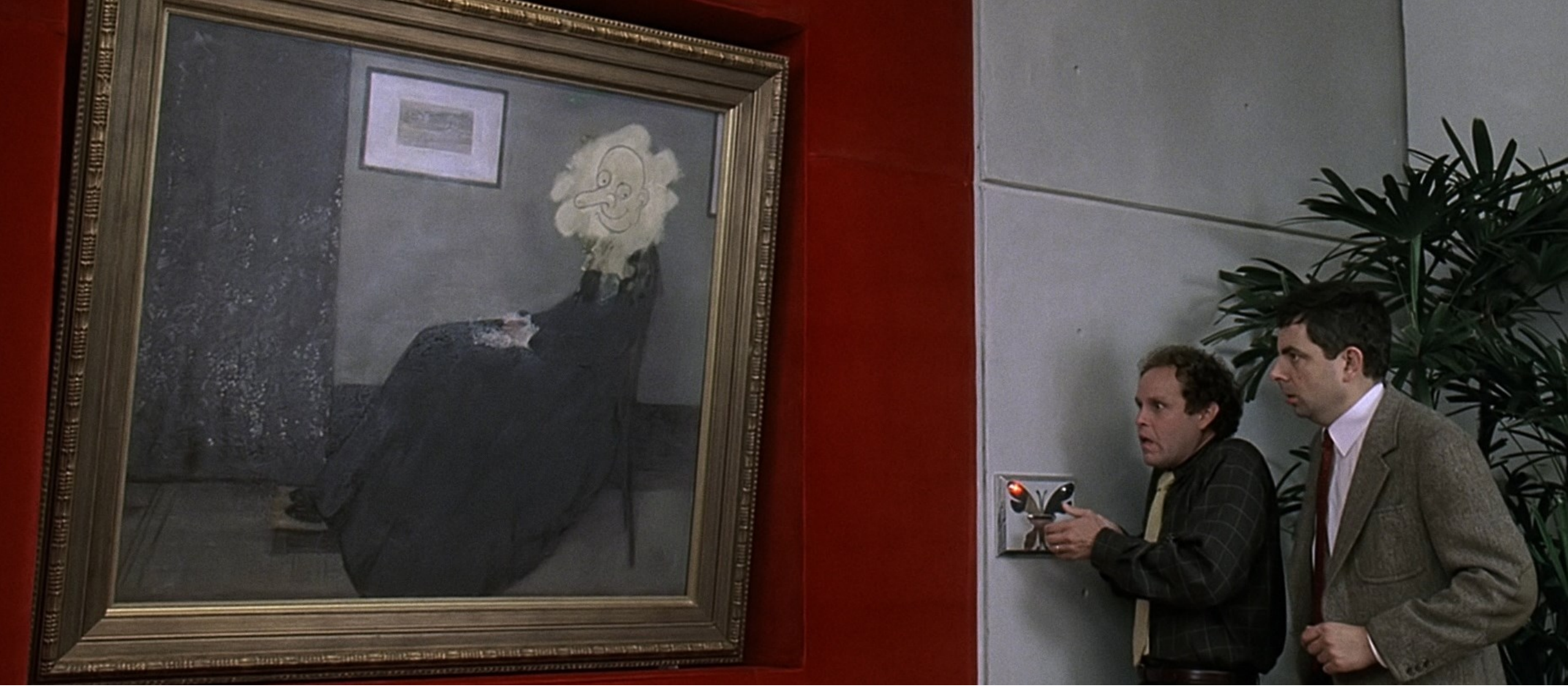


Scene from Bean: The Movie

Profound or *Mundane*?

Analysing profoundness through NLP

By: Feng Weihan

Profound or *Mundane?*

Too often we enjoy the comfort of opinion without the discomfort of thought.

Profound or *Mundane?*

Like most men my age, I'm 51.

Profound or *Mundane?*

Too often we enjoy the comfort of opinion without the discomfort of thought.

John. F. Kennedy

Profound or *Mundane?*

Like most men my age, I'm 51.

- u/awesome_smokey

Problem Statement

profound adjective

pro·found | \ prə-ˈfau̇nd, prō- \

Definition of *profound* (Entry 1 of 2)

1 a : having intellectual depth and insight

b : difficult to fathom or understand

1. Is there a pattern to profound sentence?

2. Can we develop a reliable subreddit classifier to classify statements from r/oneliners vs r/quotes?

3. Can the classifier generalize to identify any profound statement?

This project seeks to analyse patterns of profound sentence, and develop a generalized classifier to recognise a profound statement, in order to enhance our understanding of this deeper end of the human ingenuity.

Subreddits

Profound

29,718 (57%) posts from

r/quotes

Current quotes, historic quotes, movie quotes, song lyric quotes, game quotes, book quotes, tv quotes or just your own personal gem of wisdom.

Mundane

22,479 (43%) posts from

r/oneliners

A oneliner is a succinct, funny or witty remark. The statement should fit into one sentence.

Features Used

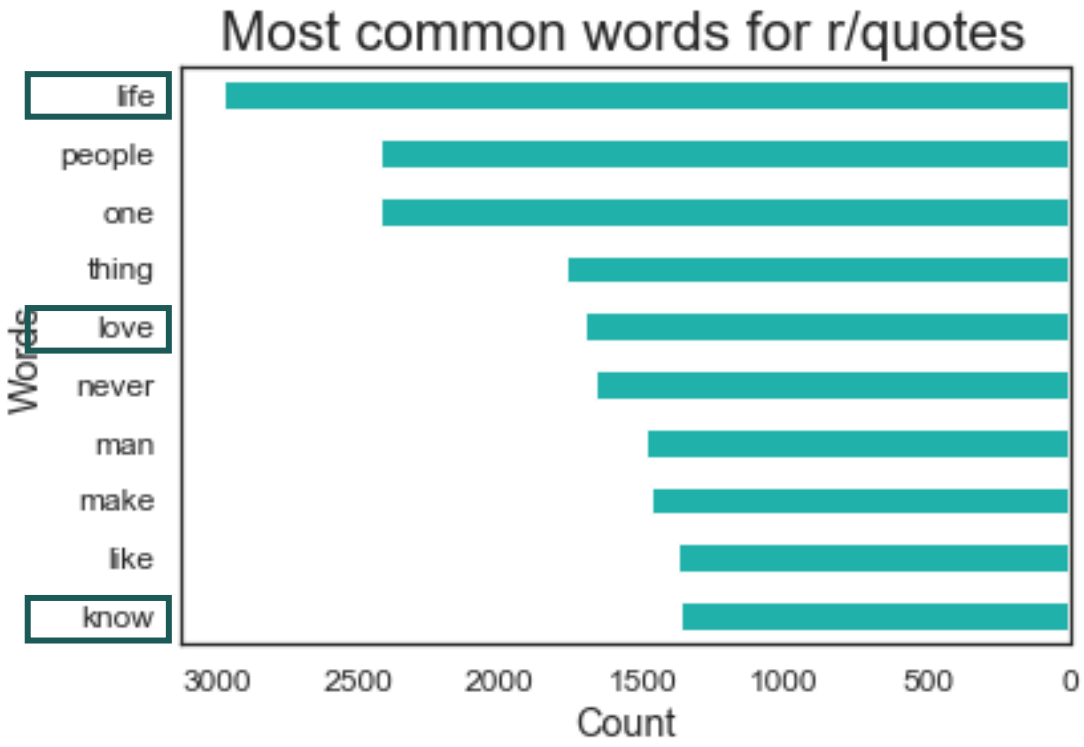
Features

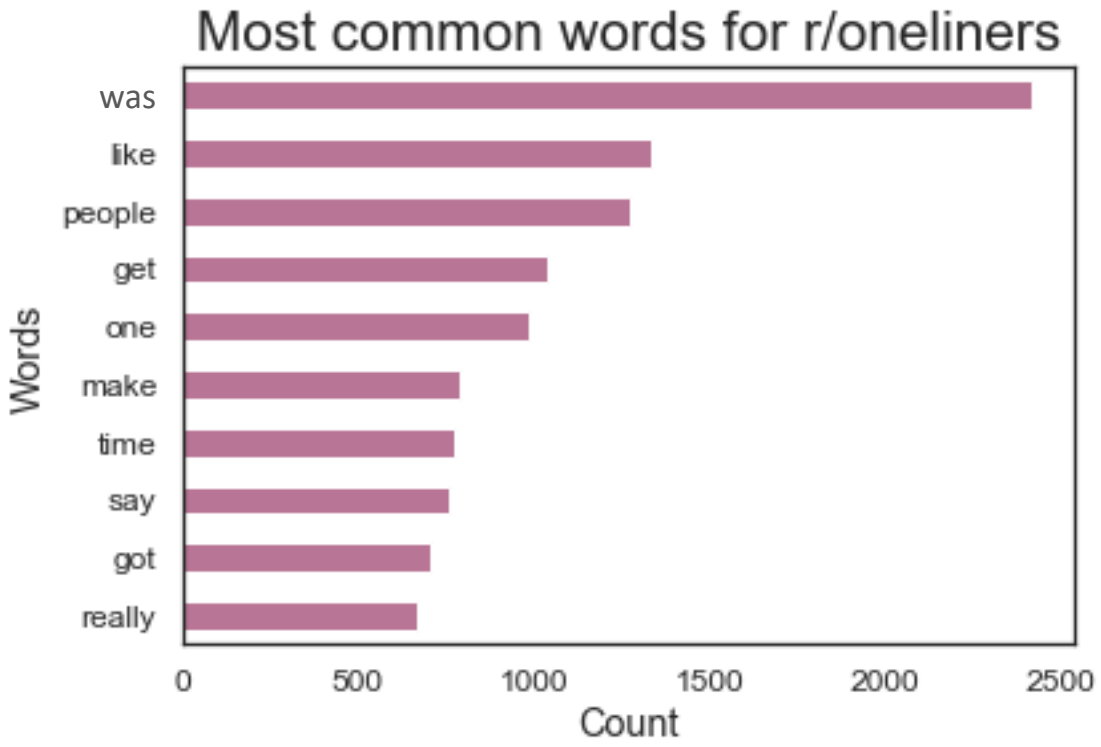
- Post Titles
- Title Length
- Average Word Length
- Longest Word Length
- Dale-Chall Readability Score
- Flesch–Kincaid Reading Ease
- Sentiment Score (VADER)
- Part of Speech (POS) Tagging

- Comments
- Number of comments
- Date posted

1. Is there a pattern to profound sentence?

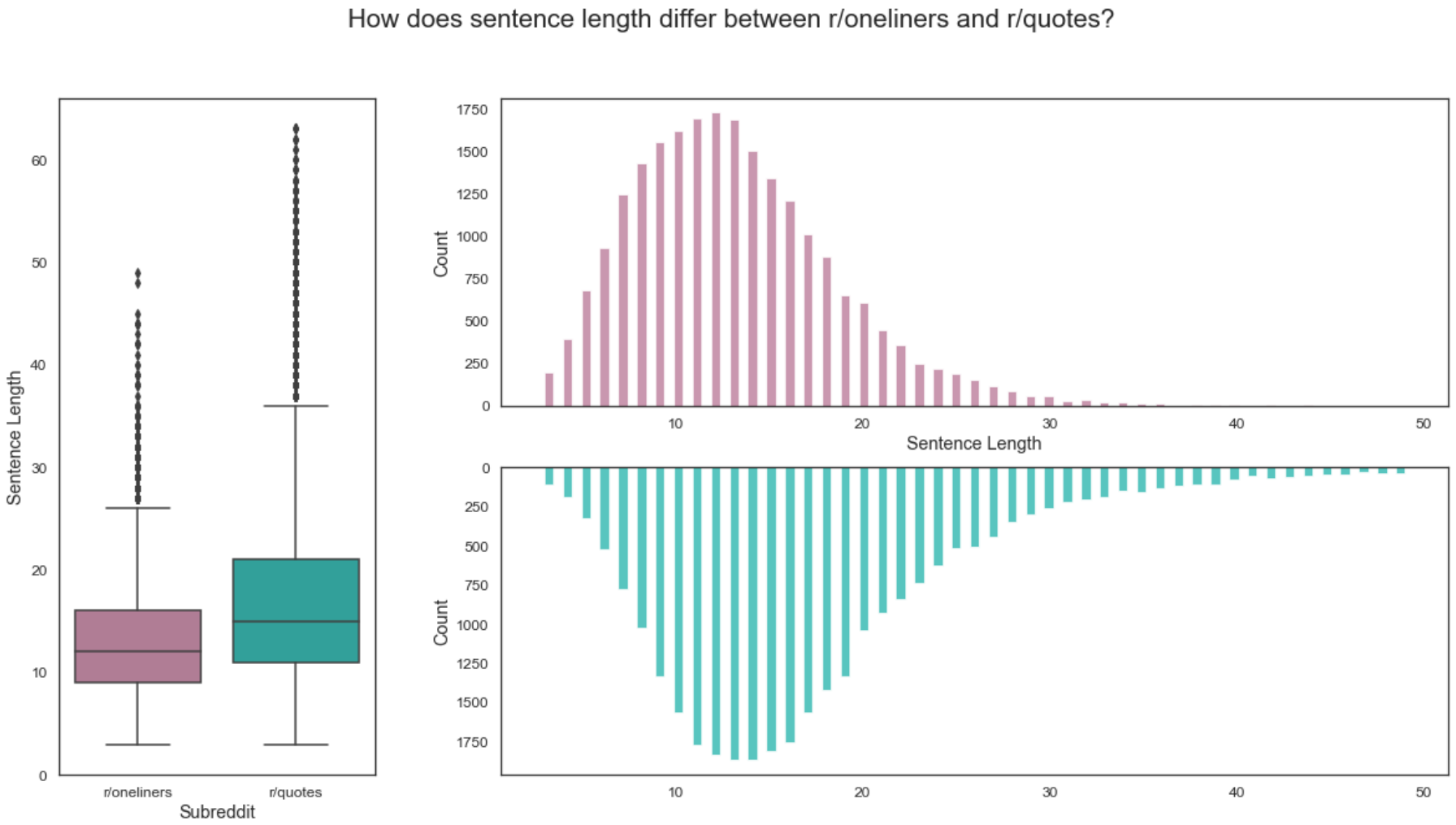
Most Common Words





1. Is there a pattern to profound sentence?

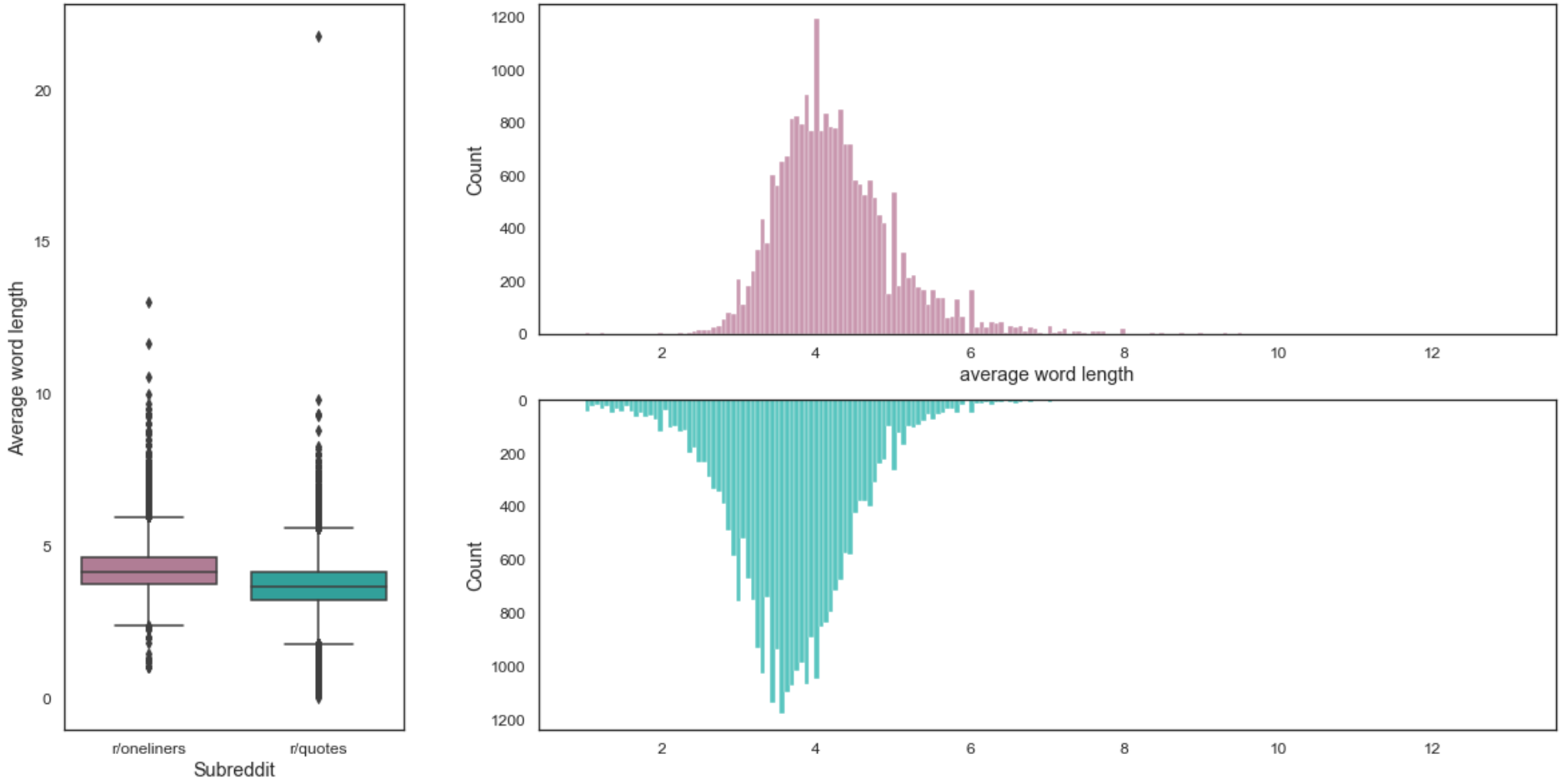
Sentence Length



1. Is there a pattern to profound sentence?

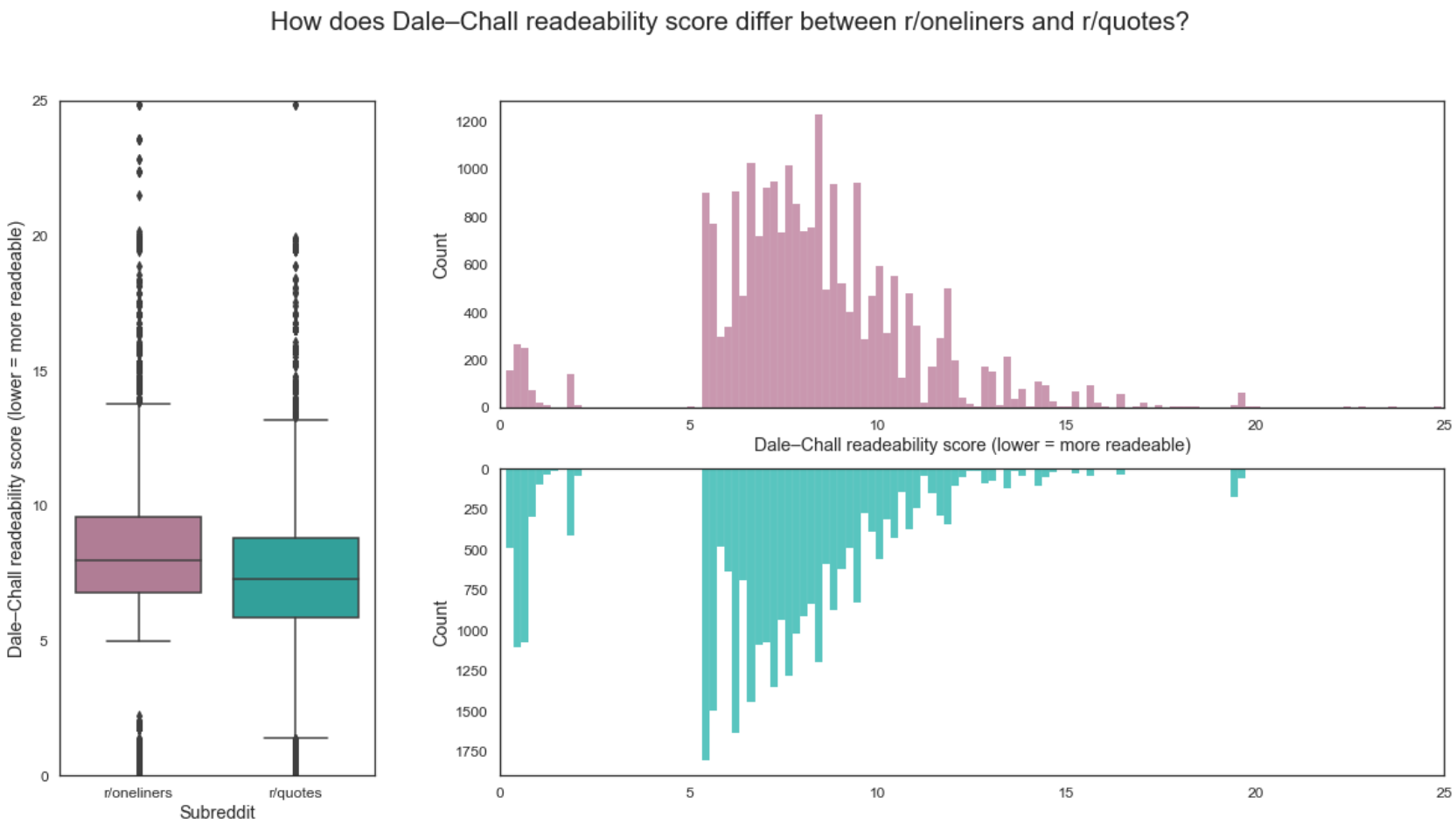
Average Word Length

How does average word length differ between r/oneliners and r/quotes?



1. Is there a pattern to profound sentence?

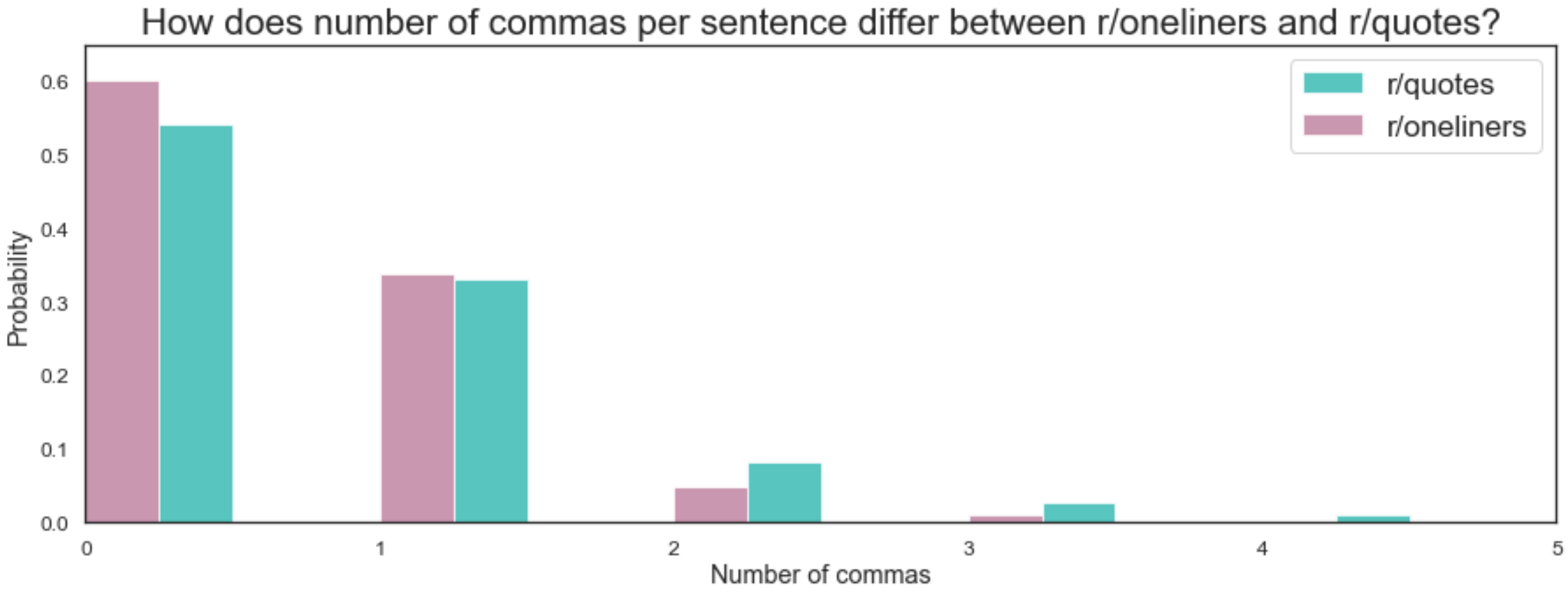
Readability - Dale–Chall Readability Score



- Function of percentage of difficult words and word length
- Lower = More Readable

1. Is there a pattern to profound sentence?

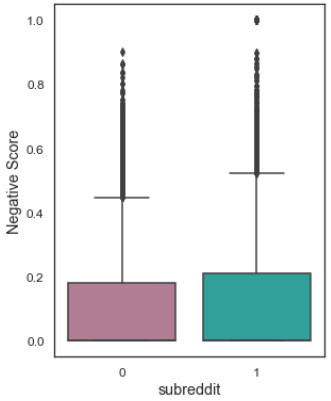
Number of commas

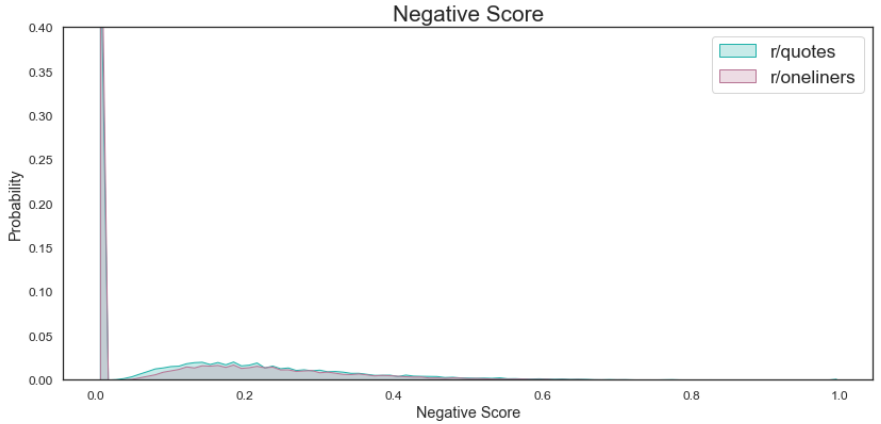


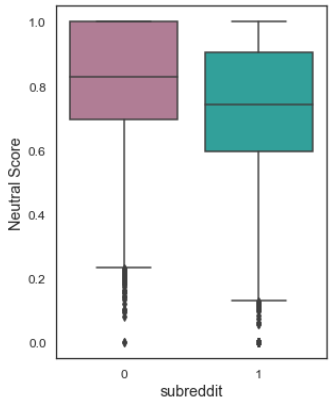
- r/quotes more commas

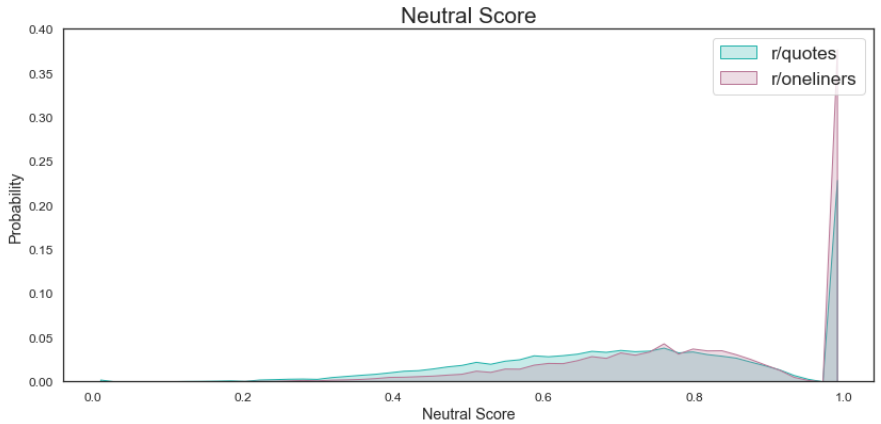
1. Is there a pattern to profound sentence?

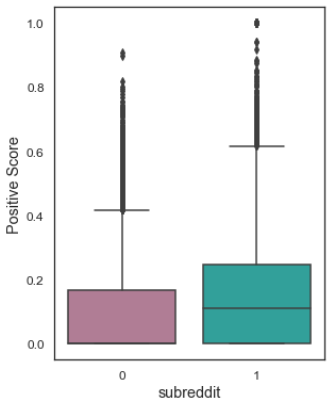
Sentiments

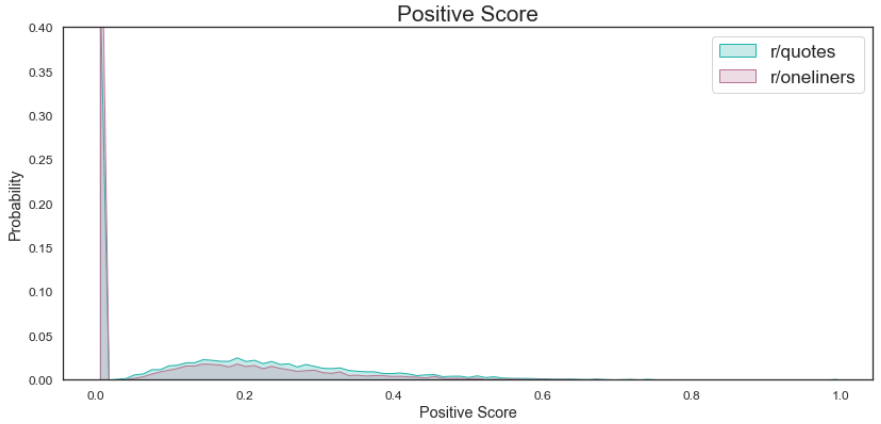


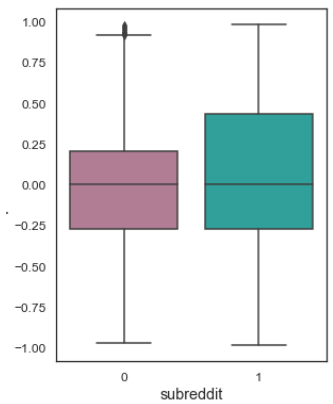


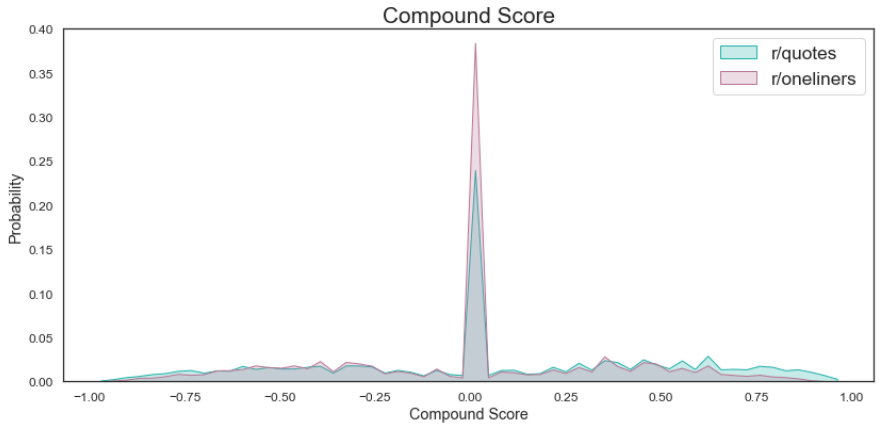












- r/quotes more positive, r/oneliners more neutral.

1. Is there a pattern to profound sentence?

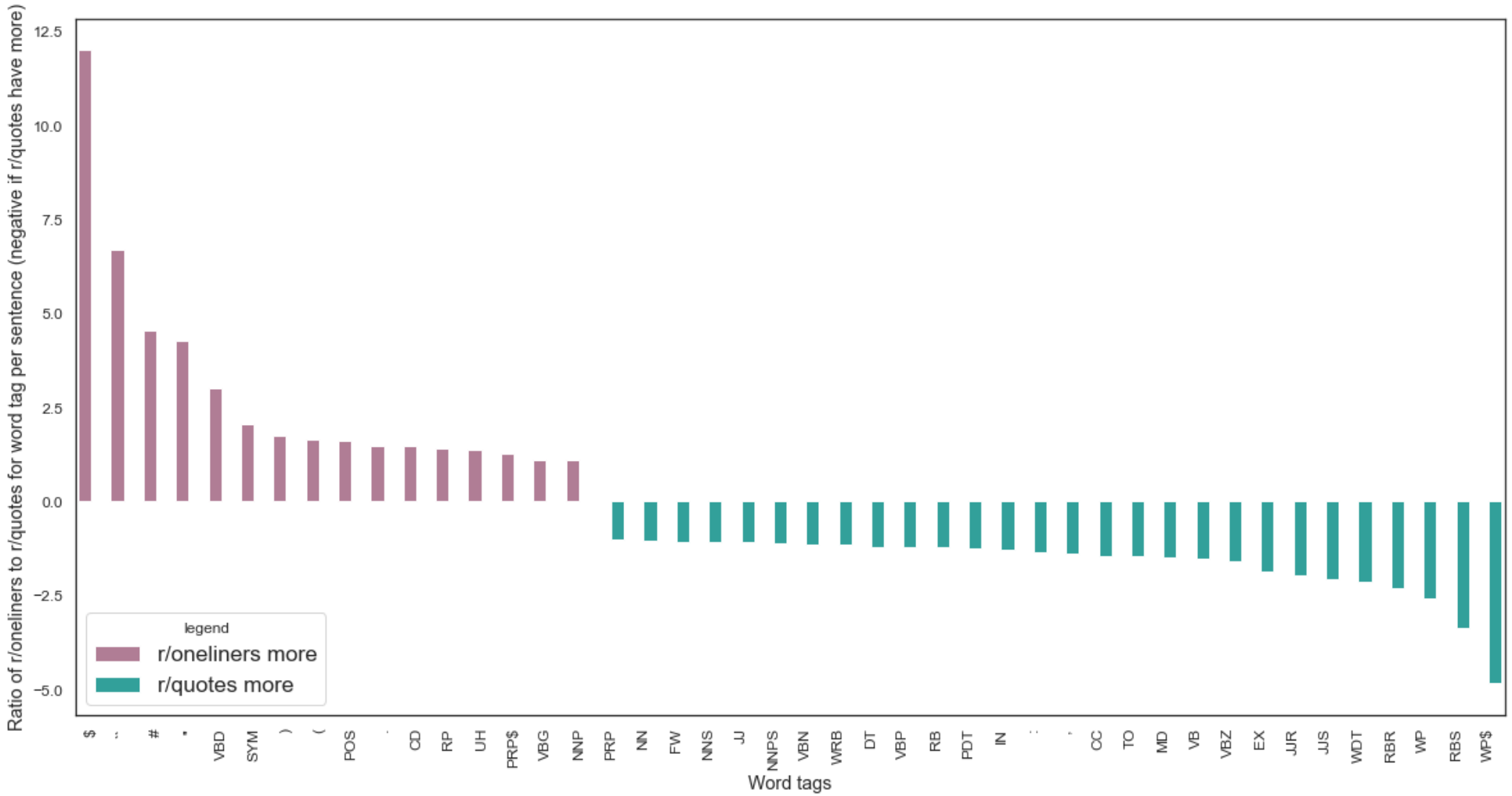
Part of Speech (POS) Tagging

Adding tag/label to each word of a sentence.

Why	not	tell	someone	?
adverb	adverb	verb	noun	punctuation mark, sentence closer

1. Is there a pattern to profound sentence?

Part of Speech (POS) Tagging



1. Is there a pattern to profound sentence?

Part of Speech (POS) Tagging

r/quotes

- Tags which increase sentence complexity
- such as pronoun(i.e. *whose, whom*), conjunction(i.e. *and, nor*), and comparative adjective(i.e. *best, bleaker*)

r/oneliners

- Tags which makes a sentence less “formal”
- symbols(i.e. :,!) , numerals(i.e. 1,2,3), proper nouns (i.e. Escobar, Liverpool), as well as verbs with a time setting, such as past tense (i.e. *made*), and present participle (i.e. *making*) , as well as interjection like (i.e. *oops, huh*)

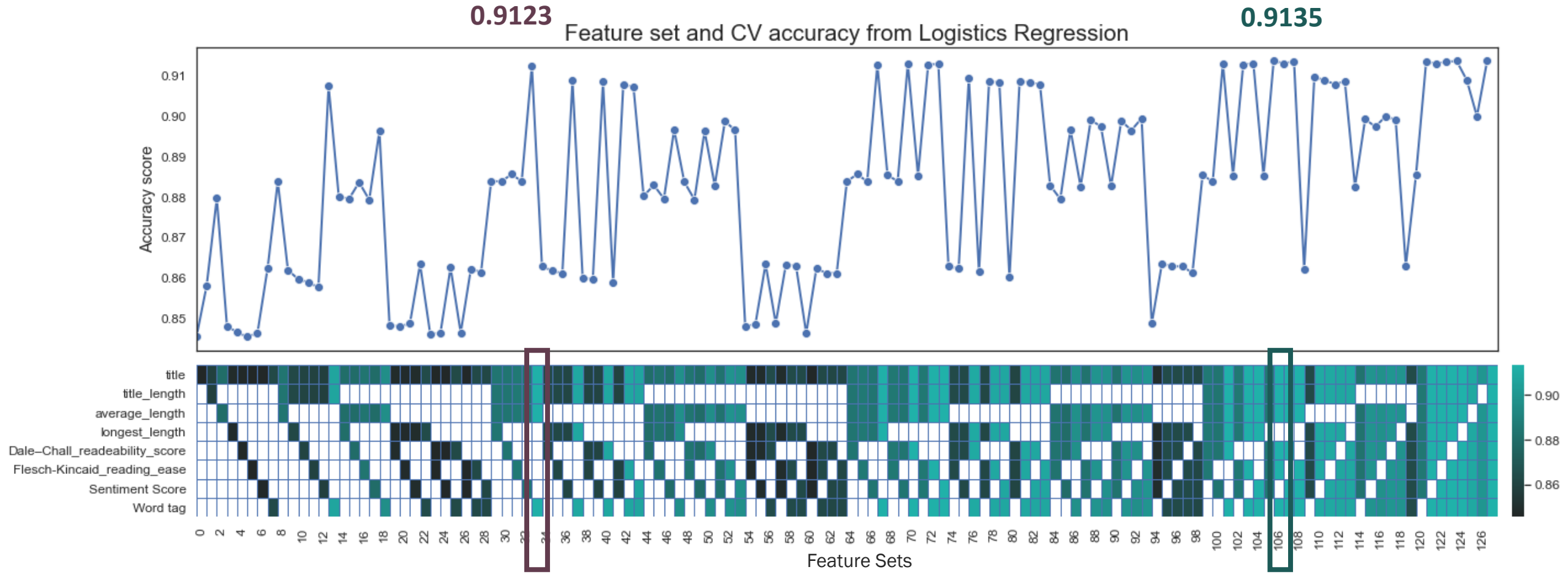
2. Can we develop a reliable subreddit classifier ?

Features

- Post Titles
- Title Length
- Average Word Length
- Longest Word Length
- Dale-Chall Readability Score
- Flesch–Kincaid Reading Ease
- Sentiment Score (VADER)
- Part of Speech (POS) Tagging

2. Can we develop a reliable subreddit classifier ?

Features



Selected: Feature Set 33

Best: Feature Set 106

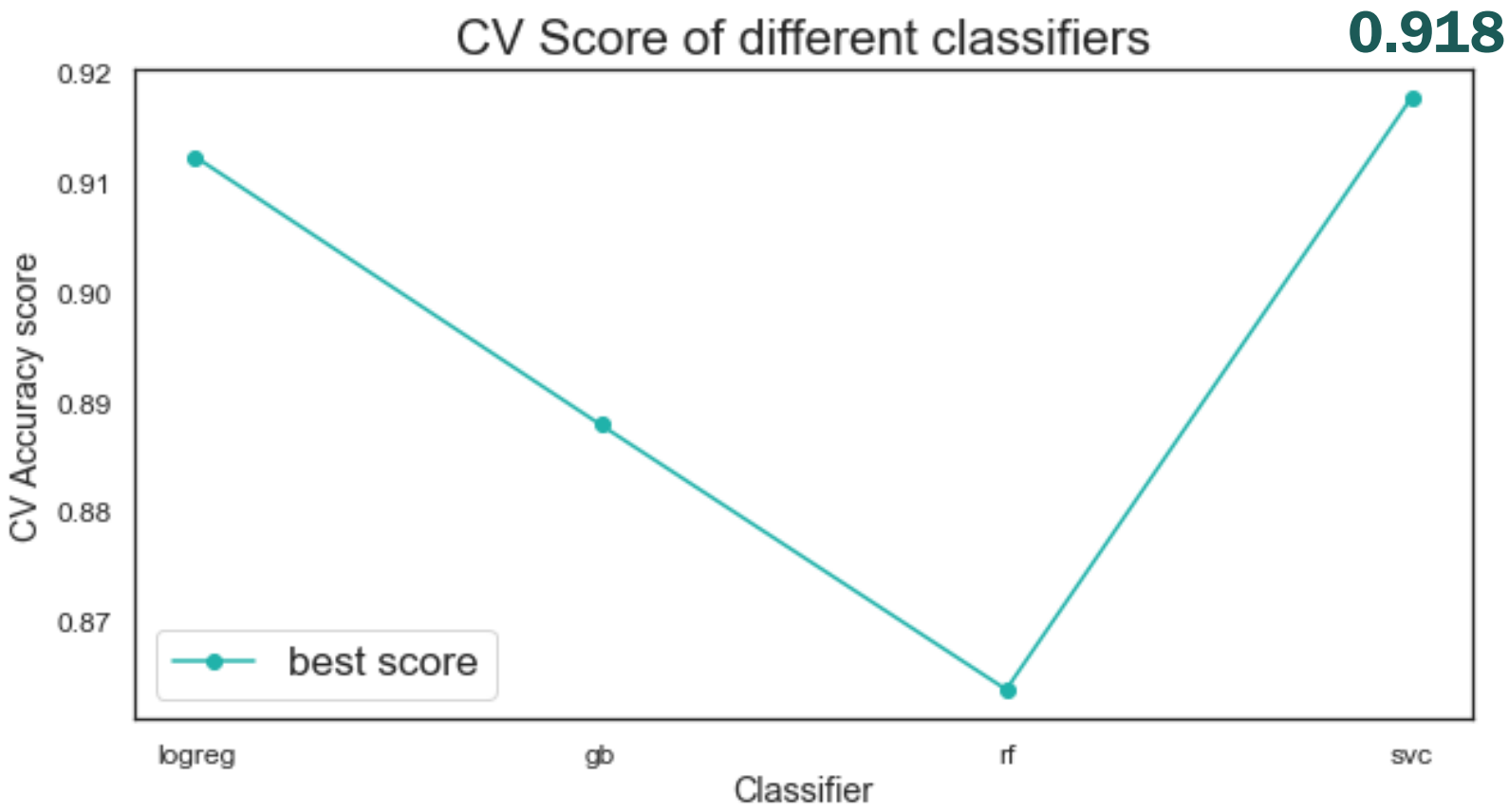
2. Can we develop a reliable subreddit classifier ?

Features

- Post Titles
- Title Length
- Average Word Length
- Longest Word Length
- Dale-Chall Readability Score
- Flesch–Kincaid Reading Ease
- Sentiment Score (VADER)
- Part of Speech (POS) Tagging

2. Can we develop a reliable subreddit classifier ?

Classifier Comparison

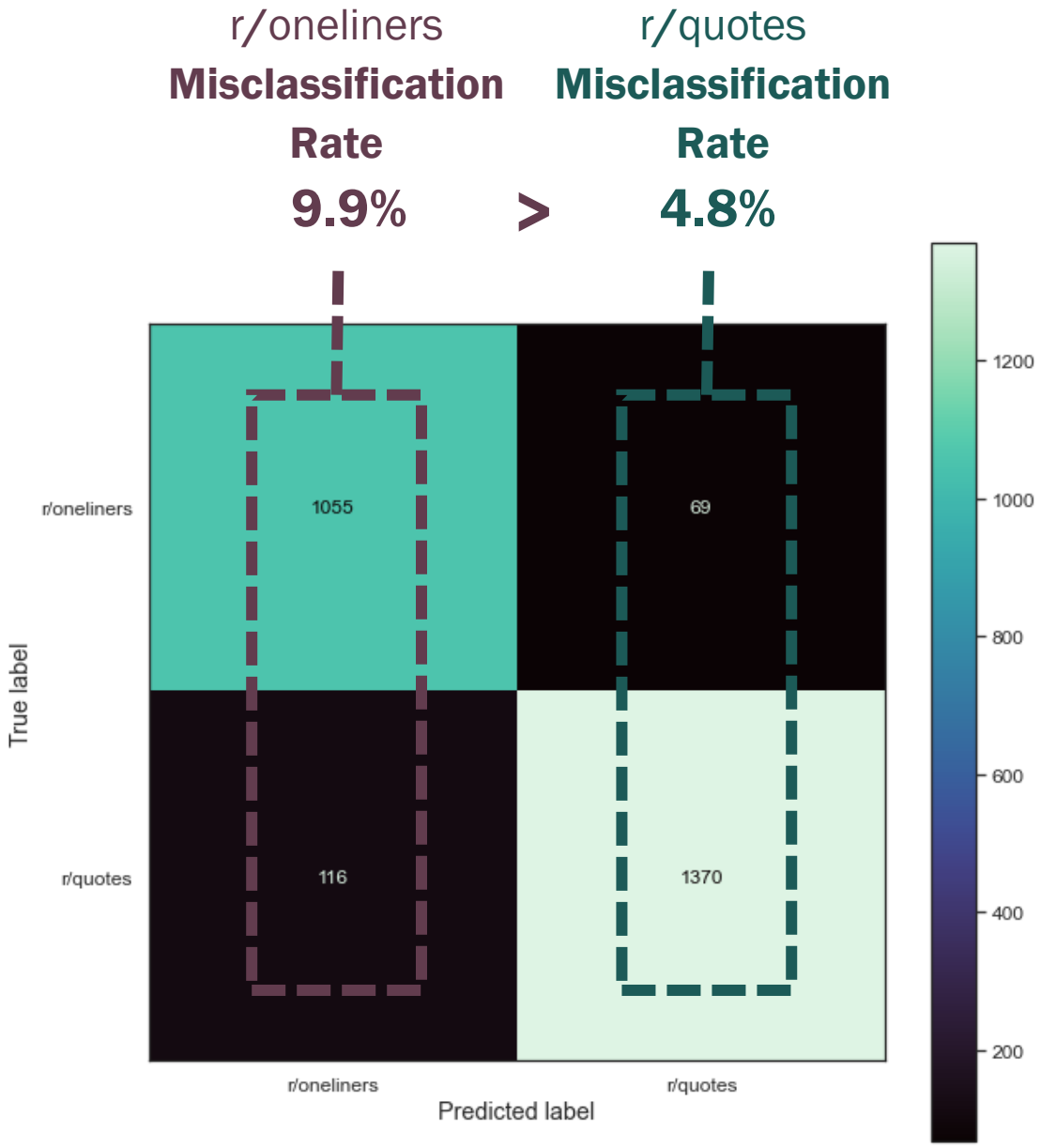


0.918

2. Can we develop a reliable subreddit classifier ?

Evaluation (Test data)

- r/oneliners has a higher misclassification rate
- Could be due to r/oneliners containing profound quotes too



2. Can we develop a reliable subreddit classifier ?

Evaluation (Test data)

Sample misclassifications from

r/quotes

- Life is like a fan, it keeps moving till someone shuts it off
- Stalking is when two people go on a long romantic walk together but only one of them knows about it.
- Well, i guess i'll have my flesh be replaced with bugs
- A thing is a thing not what is said of that thing.

r/oneliners

- Authority will always resort to violence if ignored.
- Your past and future can deceive you, but your present is always in the know
- Golf is the only sport where the purpose of the game to to play less of it.
- Duct tape is like the force, it has a light side and a dark side and it holds the universe together.

3. Can the classifier generalize to identify any profound statement?

Evaluation (Twitter data) - generalizability

From @b0ringtweets – Very very mundane texts

Samples

- Retweet if you're sitting down.
- I've been asleep but now I'm awake.
- A man is being made the President.
- Steve just texted back.

Accuracy = 98%

128/131

3. Can the classifier generalize to identify any profound statement?

Hands On Evaluation

Conclusion

1. ***Is there a pattern to profound sentence?***

   Yes, such as sentence length, word length, word tags

2. ***Can we develop a subreddit classifier to reliably classify statements from r/oneliners vs r/quotes?***

   Yes, the classifier did rather well in terms of classifying subreddits.

3. ***Can the classifier generalize to identify any profound statement?***

   Yes, to texts which are obviously not profound.

   BUT…Can be quite easily fooled by a human.

Limitations and Future Directions

1. ***Cleaner data – samples which are correctly labelled***

2. ***Word corpus – a corpus of profound terms or shallow words/internet slangs as reference for our classifier***

3. ***Contextual analysis – Methods to understand context and relations***