
NLP project

— Sentiment analysis on death
penalty in Singapore —

Members:
Sagaran

Background

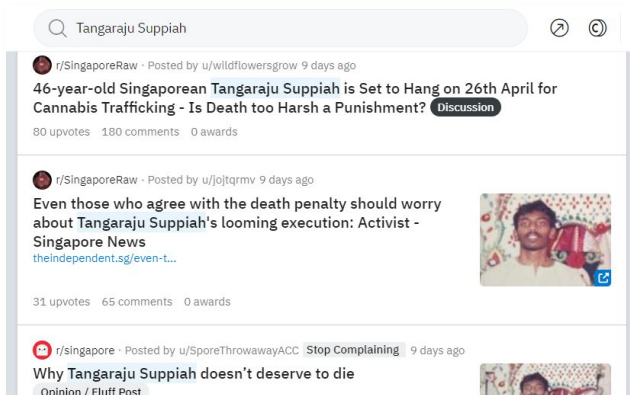
- Singapore has a tough stance on drug trafficking and drug consumption [1]
 - Intentional murder and trafficking of significant quantities of drugs could lead to capital punishment
- A recent death penalty on a drug trafficking case brought international attention on Singapore's strict judicial processes
 - It stimulated discussions within the island and across the globe alike
 - Western media and high profile members expressed their disapproval towards this approach and labelled it as draconian measures
- Singapore government gauges public opinion on this topic frequently [2]
 - The latest study/ survey predates the latest high profile individual, Mr. Tangaraju Suppiah, 46 (RIP)
 - Therefore, it is important to study the latest public sentiment on this ever changing landscape

Problem statement

- Research question:
 - What is the public opinion on death penalty in Singapore based on the recent execution?
- Success criteria:
 - Classify public sentiment on this topic – Positive, Negative and Neutral
 - Focus analysis on local Singapore sentiments on the topic
- Stakeholders:
 - Primary stakeholders: Singapore public sector policy makers
 - Secondary stakeholders: Informed citizens of Singapore

Data collection

- Data is extracted from reddit using API (praw package)
- Data collection is focused in the Singapore context by restricting the search to relevant subreddits
 - subreddits = ['Singapore', 'SingaporeRaw', 'asksingapore', 'trees']
- Extracted comments were organized into a DataFrame for analysis



Column	Description
date	Date of comment/ reply
sub_id	ID of the subreddit
title	title of the subreddit
comment_id	ID of the comment
comment_author	Author of the comment
comment_text	Main text of the comment
upvote_score	Upvote score from the members
comment_or_reply	Identifier for comment or reply
sentiment	Sentiment of the comment_text using Vader pretrained model

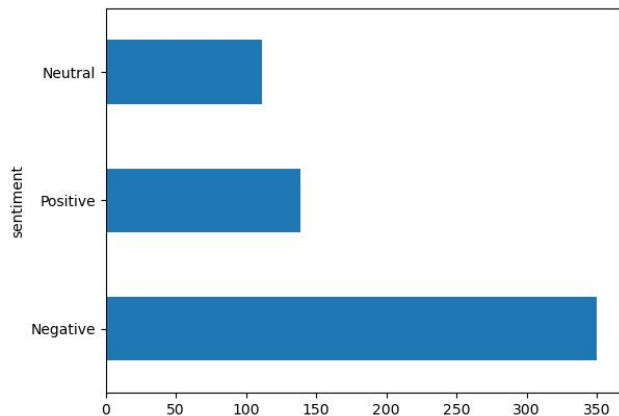
<https://towardsdatascience.com/scraping-reddit-data-1c0af3040768>

[https://www.reddit.com/r/redditdev/comments/5hzkpr/using praw to search posts on reddit works for](https://www.reddit.com/r/redditdev/comments/5hzkpr/using_praw_to_search_posts_on_reddit_works_for)

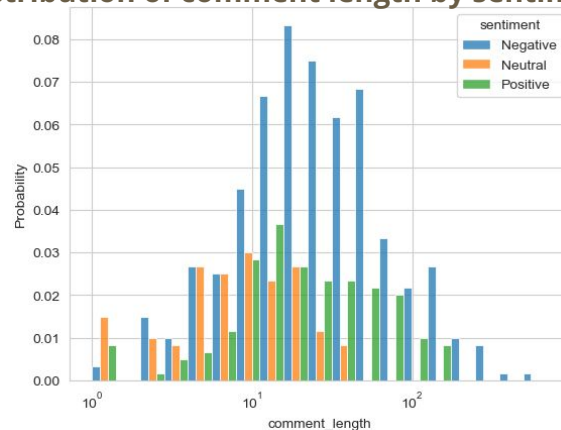
EDA

- Majority of comments had a negative sentiment.
- The data range of the reddit posts was 22 Apr to 26 Apr when this topic was hot.
- The sentiment showed a correlation on the comments length, ie. longer comments tend to show negative comments and shorter comments tend to be neutral comments

Distribution of extracted sentiment



Distribution of comment length by sentiment



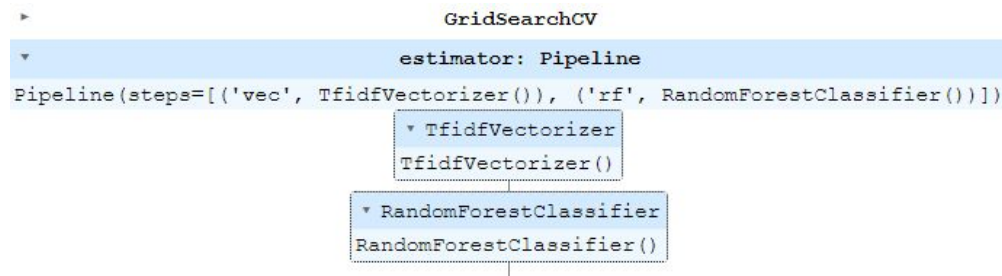
Preprocessing data

The following preprocessing steps were applied to make the pretext comments suitable for modelling.

- Comments that were '[deleted]' and '[removed]' were excluded
- Regex was used to remove punctuations and hyperlinks
- Standard NLP data preparation steps were applied:
 - Tokenization and removal of stop words
 - Lemmatizing
 - Stemming

Modelling

- The following models were explored to predict the sentiment.
 - Naive Bayes
 - Random Forest
 - KNNNeighbors
- Grid search was used for hyperparameter tuning and model selection. Best model was chosen using AUC ROC metric as the sentiment distribution was imbalanced.
- Random Forest algorithm showed the best overall AUC ROC.

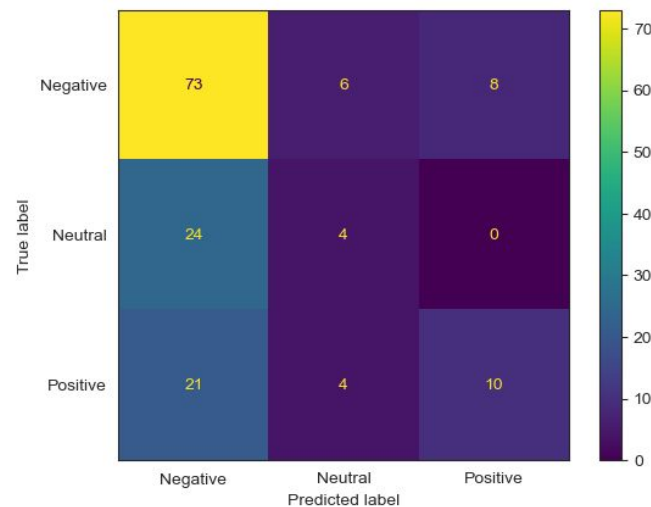


Algorithm	AUC ROC
Naive Bayes	0.73
Random Forest	0.75
KNNNeighbors	0.66

Model Metrics

- The final random forest model was evaluated on the test set to calculate the model performance metrics.

	precision	recall	f1-score	support
Negative	0.618644	0.83908	0.712195	87
Neutral	0.285714	0.142857	0.190476	28
Positive	0.555556	0.285714	0.377358	35
accuracy	0.58	0.58	0.58	0.58
macro avg	0.486638	0.422551	0.426677	150
weighted avg	0.541777	0.58	0.536679	150



Conclusion and next steps

- More than 50% the comments exhibited negative sentiments towards the death penalty
- Our multiclass machine learning model had a good ability to identify the negative sentiments, however, needs further improvement to enhance its performance