Module Name: Advanced Crime Analysis
Module Code: SECU0050
Candidate number: ZWNF4
Word count:1999
Date: 16/04/2019

Introduction

Hate crime in the virtual space is a common occurrence, from video game chat forums to specific forums dedicated in expressing hateful views. One such website is 4chan. 4chan is a collection of online forums where anonymous users can post comments and images on specific topics. Topics can range from hobbies such as fishing, gaming and cartoons to more nefarious subjects such as rape and hate crime (Dewey, 2014). In addition, 4chan acts as an haven for problematic/dangerous individuals to speak their mind. For example, the Virginia Mall and Oregon College Shooters may have posted about their crime before the incident (Kravets, 2013; Marren, 2015). Therefore, efforts must be placed on threads about violence as one of these individual may be a potential offender.

Aim

With the recent New Zealand shooting, there have been several threads opened in 4chan discussing the incident. In fact, internet services in New Zealand even temporarily blocked access to 4chan. This project will focus on one of these threads. The aim is to carry out a sentiment analysis on the comments and to create random forest classifier using supervised machine learning. The model will be used to predict if comments are “For” or “Against” the shooter’s beliefs as sympathizers of the attack may be dangerous.

Data

The data was collected from 4chan’s politically incorrect (/pol/) threads (forums) list. The topic is about the New Zealand Mosque Shooting by Brenton Tarrant. The thread was started on the 8th of April, 2019 by a user who was against Tarrant's beliefs and actions. It is unknown whether the user was trying to gauge the views of the other users. The thread had been archived at the time of data collection. It consist of 307 posts (including the original poster) about the recent New Zealand Shooting with 23 of these post being only images. All the comments were in English. However, the language used in the discussion is mired in complexity, words sometimes referring to other memes or historical events. It is a whole new subcultural language created by the users. In some cases words are broken down to near incomprehensible gibberish. For example, Turkish people are referred to as Turkroach or just roach (a combination of Turkish and cockroach). Another case would be how Muslims are referred to as Mudslimes. On the other hand, there were users constructing conspiracies to explain the shooting, all of which placing the blame on ethnic Jews by mentioning Mosad. Mosad referring to the Israeli Intelligence community. The complexity of the language and theme made it difficult to completely categorize the comments as “for” or “against” the shooting.

The comments were classified as either “for” or “against”. This was done by manually as the complexity of event requires knowledge of past events such as the Rotherham child sexual exploitation scandal and Brenton’s manifesto. Comments “For” the shooting needed to align with the ideals expressed by the shooter. Simply, they had to be anti-immigration, in favour of the shooting, praise the shooter, and support an all-white Europe. The comments did not need to meet all the criteria above. As long as they met one or a combination, they classified as "For". Moreover, it is important to mention that comments that dislike the shooter but agreed with his belief were classified as “for”. It is also worth mentioning that tread was mixture of the supporters, opponents, and general trolls (people making sarcastic/random statements). Therefore, comments that are against the shooter comprise of opponents and just random trolls.

Method and Analytical Plans

Firstly, web scraping was used to collect the comments of the thread from 4chan. The text data was turned into a csv file. Secondly, Numeric values were removed and a sentiment analysis was carried out on the comment section using the sentimentr package, both for individual sentences in each post and the whole post.

Afterwards, the dataset was cleaned by removing digits. Descriptive statistics were carried out on the following components 1) number of sentences, 2) number of tokens, 3) number of types, and 4) type/token ratio. The digits were removed before descriptive statistics because the number represent users IDs and formed a single token with the first word of the comment.

Third, the data was partitioned into test set and training set, divided into 30% and 70% respectively. Both the These dataset were tokenized with the following elements removed; 1) numbers, 2) punctuation, 3) stem words and 4) hyphens. The dataset for here changed into a document frequency matrix and then into term frequency inverse document frequency (TFIDF) matrix using the quanteda package. The same processes were used to create the trigram TFIDF and bigram TFIDF.

Forth, cross validation was carried out on following datasets; 1) TFIDF, 2) trigram TFIDF and, 3) bigram TFIDF. The datasets were split into 10 k equal partitions and the random forest algorithm was used with 3 repetition (to save on computational power). Lastly, the testing set was used to check the accuracy of the model. For the classification, random forest algorithm was used because it is an accurate predictor, the classifier will not overfit the model as long as there are enough trees, and it is robust in handling missing data.

Results

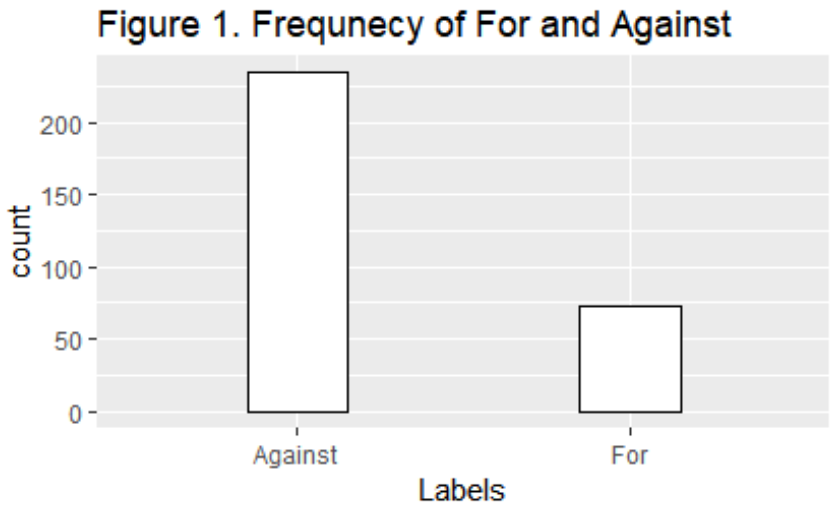


Figure 1 shows the number of comments "For" and "Against" the shooter. The percentage of "Against" comments are nearly three times of the comments who are "For" (Against=76% and For=23%). This was surprising as the data was collected from the politically incorrect forum which should have had more extreme right leaning individuals.

The average number of sentences was 1.8 with a maximum of 19 and a minimum of 0. Due to the minimum number of sentences being 0, the minimum of token and type were not calculated. The average number of tokens 32.67 with a maximum of 440. On the other hand, the average number of types is 24.92 and the maximum is 229. The mean type/token being 0.77, meaning near all the words used were unique.

Figure 2.1. Histogram of frequency for Types

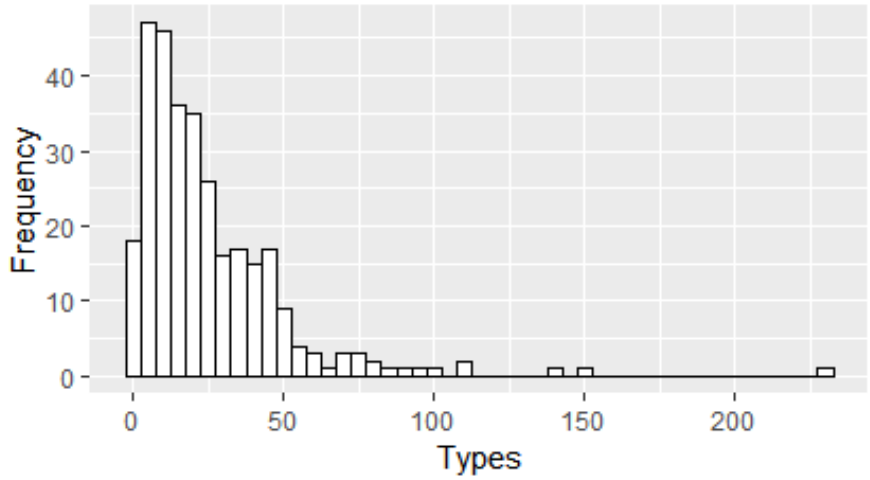
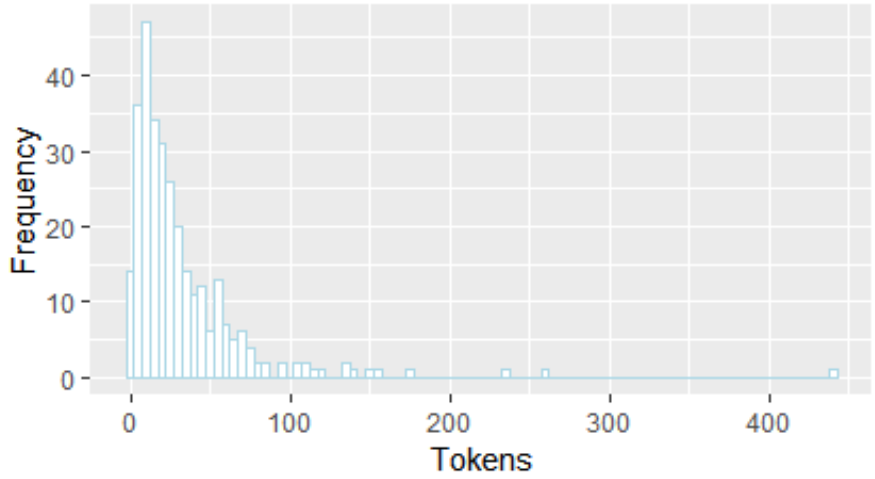
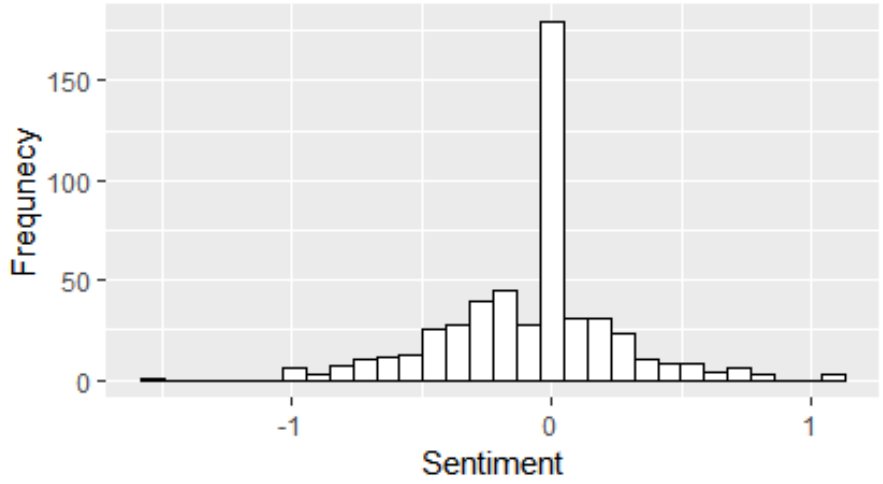


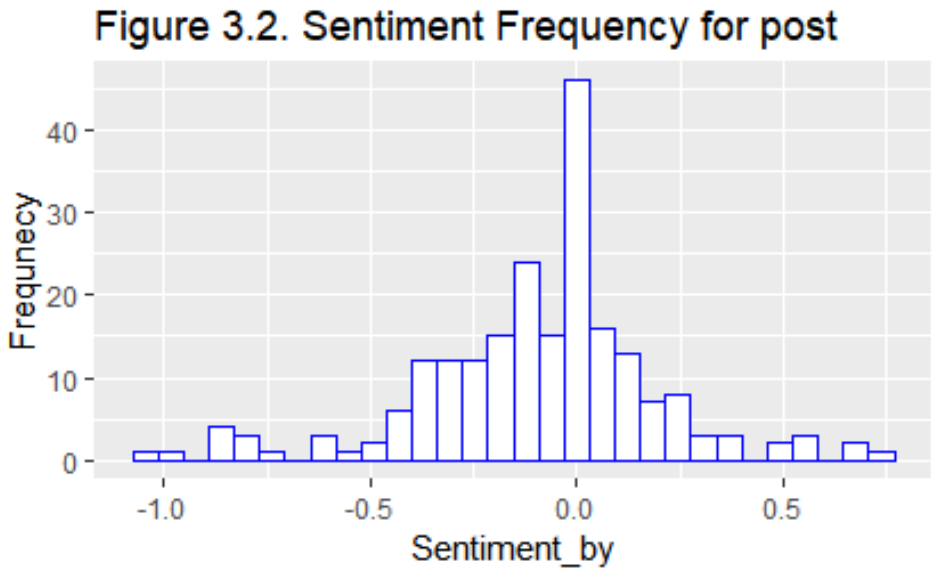
Figure 2.2. Histogram of frequency for tokens



Both Figures 2.1 & 2.2 show that the type and token distribution follow Zipf's law of language (Zipf, 1949). Excluding the posts that have no text, the histogram follow a negative exponential decay curve.

Figure 3.1. Sentiment Frequency for individual





Figures 3.1 & 3.2 show that most of the posts and sentences in the posts are neutral. In general, sentiment tills towards the negative spectrum. However, sentiment analysis is not reliable in indicating whether the negative sentiment is directed towards the shooter or directed at the victims.

Training set results

The most accurate results were from the TFIDF model, followed by the bigram, while trigram being the least accurate. For the TFIDF model the highest accuracy was 77% from 1388 branches, Kappa=0.25, and the lowest was 74% from 52 branches, kappa=0.02. On the other hand, the best score for the bigram matrix was 76%, 2746 branches, Kappa=0.06. In contrast, trigram matrix had the lowest accuracy score, 75%, 2 branches, kappa=0. Results show trigram model to be the most unreliable. Due to the TFIDF model having the highest accuracy, the testing set will be tested with it. However, the bigram model shows promise and testing set will also be tested with it.

Test set results

The model created for the purpose of predicting comments, whether they are "For" or "Against" the New Zealand shooter's belief, is about 74% accurate. The model scores for Recall, Precision and F1:

1) Recall; the model can correctly predict (Sensitivity) 92% of the comments that are "Against" the shooter's belief. On the other, the model can correctly predict only 14% of the comments that are "for" the shooter's belief

2) Precision; the model has precisely predict "Against" 78% of the time, while only 37% for "For"

3) F1 score: the model scored 0.81 which is high score for Against comments. However, it score 0.2 for "For" which is low.

Results of the model on bigram are more unreliable when compared to the results of the TFIDF model. The overall accuracy of the model increased by 4% (from 74% to 78%). It scores are:
1) Precision; 70% for "Against" and 100% for "For"
2) Recall; the model has a sensitivity score of 1.00 and specifiticity score of 0.07.
3) F1 score; the model scored 0.13 for "For".
It was determined that there was no need to carry out an Area Under the Curve plot as the model was not successful in predicting the class of interest, "For". The goal of this project was predict comments that where in support of the New Zealand shooter's, however, the results shows that the models cannot estimate with significant reliability the comments of interest.

Conclusion

The result of the project shows that prediction of the supporters of the New Zealand Shooter was not possible with the current sample of comments. The models (TFIDF & Bigram) are able to classify comments that were "Against" the shooter, however, that is not the intended outcome. Comments against the shooter also included random/general comments. In other words, these comments lack any uniqueness. On the other hand, comments in support of the shooter were too diverse in their ideological and political expression. In fact, one of the post mentions 4chan to be a diverse plain of ideologies. This means that comments in support of the shooter had few similarities, therefore, the TFIDF and Ngram matrix did not show any numeric significance.

The interpretation of the sentiments of the comments was difficult. As a commenters stating they want to protect the future of their race is not a negative statement. While another mentioning that they wanted the shooter to get harmed would be a negative sentiment. In other words, sentiment analysis is too naive to properly place the sentence in the morally right area. Therefore, sentiment analysis of the comments for this sample has little meaning. In fact, the morality of the views expressed by the shooter is depended on a person's political leaning. What the shooter did is a crime but thinking of the safety of one's own race or immigration are subjects of contention. For example, the classification of someone who calls the shooter a hero is easier than someone who is anti-immigration and pro-white but makes no mention of the shooter. Therefore, the classification of comment may not be reliable due to the lack of inter-rater reliability.
In addition, the language used in these post is unique to the forums. As Davison (2012) explains, meme are more infectious than spoken language. Memes are what change the behaviour of the individual through experience. Therefore, it is likely that the nature of 4chan attracts user who are well versed in this culture of memes. There are comments referencing events from Rotherdam child grooming scandal to connecting the incident to the Israeli Intelligence community (Mossad). It is near gibberish to those who have not been part of the community for long. The cultural barrier between the user and the researcher may lead to the inconsistences with the validity of the data classification for future research. There is also the problem of typos and grammatical errors, which seem to be a common presence. The lack of proper punctuation could have interfered with sentiment analysis. Furthermore, typos of words may have shifted the TFIDF score of the words. In addition, the multiple spelling of words could lead to the TFIDF scores being irrelevant.

Future research may want to focus into the actual lingo used by the users. The language used in such online forum is constantly changing with time and popular culture. Another route of interest is image based analysis of memes. However, this may prove to be difficult as memes change with time and the context of the memes are not constant. In general, future research on textual data on racism, sexism and other hate crime should take the opportunity to collect data on 4chan.

Reference:

Davison, P. (2012). The Language of Internet Memes. NYU Press, PP.120-134. Retrieved from: https://www.researchgate.net/publication/263564286_The_Language_of_Internet_Memes

Dewey, C. (2014, 25 September). Absolutely everything you need to know to understand 4chan, the Internet's own bogeyman. *The Washington Post*. Retrieved from: https://www.washingtonpost.com/news/the-intersect/wp/2014/09/25/absolutely-everything-you-need-to-know-to-understand-4chan-the-internets-own-bogeyman/?utm_term=.8edb9bd31b4f

Harris, S. (2019, 16 March). The Great Replacement: The Manifestoc of Brenton Tarrant- The New Zealand Mosque Shooter. European Freedom. Retrieved from: https://www.europeanfreedom.com/2019/03/16/the-great-replacement-the-manifesto-of-brenton-tarrant-the-new-zealand-mosque-shooter/

Marans, D. (2015, 1 October). Did The Oregon Shooter Warn of His Plans On 4chan. *HuffPost*. Retrieved from: https://www.huffingtonpost.co.uk/entry/oregon-shooter-plans-4chan_n_560da551e4b0768127016099?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAG8BWpK7chisE-oMzxFc9ETWc2gI6ql7YArZLNUcm_rDFD4rhxrV_Ca7q6Dk7MRlO5mIrqQ-sRJJ3tJ0tediGubUwMarSh9BGrMxGBzb4fvWjjgmsZFPWmkJseRUW4z7_NZMikVwWJyKMjzKG4vAlebIt1LVZmXgDAm_yLGGw_Ma

Zipf, G.K. (1949). Human Behavior and The Principle Of Least Effort: An Introduction to Human Ecology. Massachusetts, USA; Addison-Wesley Press, INC.