## 

Twitter Poets:

Sonnets on 2020 Elections

By,

Anaqi Pek 1003312

Wan Kwong Chen Ian 1003505

## 

## 

## 

## 

## 

## 

## 

## 

## On Data

By PoeTryMe

the reluctant packs of galaxy

i could do packets my data

thy proud parcels and sheaves high

my heap or bundle thereby

## 

## 

## Abstract:

Digital humanities have used digital processes such as group sentiment analysis to provide additional insight in the humanistic analysis of works. However, with the advent of generative text, one could form humanistic text from data. This paper aims to explore how humanistic analysis of generated text from big data could be used as a method of analysis, specifically through the use of poetry generation. In this paper, analysis was done on a corpus of tweets regarding the 2020 U.S. presidential elections. A Term Frequency-Inverse Document Frequency[[1]](#footnote-1)(TF-IDF) algorithm was used to identify significant words in daily tweets, which was then fed into a poetry generator. The generated poetry was then humanistically analysed to see if it could provide more insight to public sentiment in this period. It was found that the poems generated could be used to interpret certain events that happened in this duration, although due to the limitations of choosing the significant words and the poetry generator, most generated poetry did not have much relevance or meaning to be extracted from.

## 

## Introduction:

With the exponentially increasing amount of data in the world the use of big data techniques are becoming commonplace in many academic fields. Such application has seen the rise of Digital humanities, where data of the subject is created and used as a supplementary “text” in the study of the original works. One such technique is the study of group sentiment analysis, where algorithms give an estimated sentiment of the text it is given.

This is no different from the study of poetry, where humanistically analysing it can bring insight to a soft representation of ideas and perspectives it is trying to convey, albeit with a broader range of results. As such, this paper intends to explore the usage of poetry as a supplementary tool of analysis alongside group sentiment analysis in analysing a corpus.

The corpus in question is a compilation of tweets tweeted in the duration leading up to the U.S. presidential elections[[2]](#footnote-2), amounting to around 1,700,000 tweets between October 2020 and November 2020, chosen for topicality. Tweets are chosen as the social media of choice due to its relatively short text, widespread use, and responsive submissions, allowing us to get an almost real time feedback of the public emotion national events. Tweets are restricted to those with the hashtag #biden and #trump to maintain relevance. The metadata extracted from this corpus is then converted to poetry using a publicly available poetry generator PoeTryMe[[3]](#footnote-3). This creation of the poetry is further discussed under methodology.

With the data generated poetry we then use traditional techniques to analyse for meaning and apply it to the group sentiment of the tweets, obtaining new knowledge and perspectives regarding the groups represented.

## Methodology:

The data available to us was given in a csv file format which allowed us to ingest the data and perform the pre-processing needed before running any analytical tools on the corpus. Due to the raw form of the text, natural language processing was required to remove undesirable parts of the text. An example below.

before cleaning: Le président #Trump étrille Twitter et Facebook pour avoir censuré l'article du New York Post sur Hunter Biden https://t.co/iECxNDugBY

after cleaning: ### NON-ENGLISH LANGUAGE ###

before cleaning: "IS THIS WRONG??!!" Cory Booker's BRILLIANT Final Questioning of Trump Nominee Amy Coney Barrett https://t.co/gCTvVLl4CS

#AmyConeyBarrett #CoryBooker #Barrett #Booker #Trump #KamalaHarris #JoeBiden #SCOTUS #SupremeCourtConfirmation

after cleaning: IS THIS WRONG Cory Booker s BRILLIANT Final Questioning of Trump Nominee Amy Coney Barrett AmyConeyBarrett CoryBooker Barrett Booker Trump KamalaHarris JoeBiden SCOTUS SupremeCourtConfirmation

Main things to take note regarding the cleaning are the removal of the twitter image URL (https://t.co/gCTvVLl4CS), the removal of the various punctuation and the removal of non-English languages.

The data was then separated by days producing a timeline of tweets, which allowed for the use of election news and events as signposts for spotting group sentiment changes. With these signposts we were able to better pick “strong” sentiment days and focus on analysing poetry generated from those days.

The daily corpus of tweets (n =~15,000) were then put through a Term Frequency-Inverse Document Frequency[[4]](#footnote-4)(TF-IDF) algorithm and the top words were filtered out and stored. The algorithm basically takes away common words that appear in many documents. For understanding a trivial example calculation of the TF-IDF of “the”.

For 1,000 documents with total of 10,000 terms

Term Frequency = times “the” is seen/total terms=3,000/10,000

Inverse Document Frequency

= log(all documents/documents with term “the”) = log(1,000/999)

TF-IDF = (3,000/10,000) x (log(1,000/999)) = 0.3 x 0.00043 = Really Small

The small value of the TF-IDF is representative of how the word ‘the’ is lacking in meaning, hence the by only collecting words with high TF-IDF values allows for better extraction of meaning from the corpus. Resulting in a “term, score” dataset as shown below:

#biden 01/11/2020:

joebiden,0.07400749032250005

copyright,0.37039312247158607

funk,0.3986274435577389

violate,0.3986274435577389

becuz,0.3986274435577389

caps,0.3986274435577389

groove,0.3821114244890188

nation,0.23927699573156844

realdonaldtrump,0.1358449474991562

The TF-IDF values were further supplemented with the top k retweeted and liked tweets. For k=5 we extracted the top 5 retweeted tweets and top 5 liked tweets, this was then used in conjunction with the TF-IDF to produce a better word selection for the poetry generation.

These high impact words were then chosen by hand and fed to the poetry generator PoeTryMe, generating poetry from a daily corpus of tweets to surmise the sentiment of each day approaching and after the 2020 elections. The generator makes use of seed words, poem structure and a “surprise” variable which are fed into a line generator, the line generator then makes use of a semantic rules model, a grammar contextualiser and a large morphology lexicon to produce poetry[[5]](#footnote-5). In this paper however PoeTryMe is used as a “BlackBox” and further investigation and adaptation of PoeTryMe can warrant an entirely new project. More information on PoeTryMe can be found at their website <https://poetryme.dei.uc.pt/>.

PoeTryMe tends to perform better with shorter poems, this is because the generator is not perfect and tends to repeat itself, producing lines of lower quality when made to generate long form poetry. Using a poem structure of 4 lines by 7 syllables was found to be optimum.

Now armed with the poetry generated for each day tagged with either #trump or #biden we are able to map these poems to the daily election events and news to analyse and gain new perspectives on group sentiment.

## Poetry & Analysis

### #biden 25/10:

Keywords: corruption, better, definitely, confirmed

a better form with best hands

alter the door, invert me

and my build to improve free

by no infections corruption

“a better form with best hands” seems to be pointing towards a maker, perhaps a maker of opportunities where the “better form” is contextualised with the “alter the door” in line 2 which may represent how the 2020 elections were an opportunity for Biden supporters to oust Trump from presidency.

Line 2 continues with “invert me” where the maker from line one become ambiguous as if it takes on a divine position as maker and changer of the author as well as being creator of the “door”. Perhaps implying that the situation that had come to be was one of divine origin and have changed people and created a promising situation for Biden supporters. Almost alluding to the coronavirus pandemic and how daily life is now “inverted” and the “door” has been “altered” due to the Trump administration's poor handling of the pandemic.

The connection to the pandemic is further enhanced by the last line ”by no infections corruption” which takes on special meaning when Trump tested positive for coronavirus almost implying that through the trial of the pandemic Trump was proven to be corrupt.

### #biden 26/10:

Keywords: deplorable, proud, cool, love

that holds the joys of the hatred!

the love swings west, the joys follow

and deplorable and bad

from cool to coldness he had

The poem first talks about “the joys of the hatred” perhaps pointing to people on both sides of the political parties and their powerful feelings regarding the elections. The poem continues with the “the love swings west”, looking north, west is to the left this perhaps implying that the support for the left-wing party was growing and that ex-Trump supporters were now moving leftwards supporting Biden for presidency. Line 2 continues with “the joys follow” having connected joys with hatred in the first line “the joys follow” becomes a bittersweet statement that having the joys follows also include having the hatred as well. Ex-trump supporters may have generated hatred as they become seen as traitors to avid Trump supporters. That hatred which followed them to the “west”. The last 2 lines seem to point towards the presidential debate as being “deplorable and bad” while trying to paint Biden as the “cool” collected figure in the debate.

### #trump 25/10:

Keywords: die, poorer, weaker, sicker

when leaves break and cold winds die

with stamp sheets every eye

some turn live within the stopper

now they leave and die together

The first line points to spring, where leaves “break” out and the cold wind begins to “die”. Perhaps pointing to a new beginning for Trump with a potential new presidential term ahead of him. “with stamp sheets every eye” seems to imply that the viewpoints or perspectives of Trump supporters are mass produced like “stamped sheets”. The last 2 lines further elaborate on the mass produced perspectives that while quarantined “the stopper” they stay active “turn live”, causing coronavirus cases to spread “now they leave and die together”.

### #trump 26/10:

Keywords: god, time, damn

with her space and her indenture

some god live within the pantheon

earth and moon and time and moments

that it darn comes to the damn

Immediately the poem strikes with a female pronoun “her” for the hyper masculine presentation Trump supporters fabricate about Trump (see Ben Garrison) this was curious. However, the “her” may not be pointing at Trump or even Trump supporters, due to her being of divine nature, “she” potentially be what Trump supporters feel they are defending, liberty and tradition. Using this characterisation of her we see that liberty and tradition seem to be disconnected from the other gods, as she is indentured, she is contrasted with “some god(s) live within the pantheon” making her seem obligated by something that forces her to be unable to “live within the pantheon” like the other gods. The later 2 lines then seem to elaborate on her indenture, saying “earth and moon and time and moments / that it darn comes to the damn” this is also made special as the only flag on the moon is one of the USA. Almost saying that between the earth and the moon lady liberty is bound to the USA til “darn comes to the damn”, til the end of time or perhaps more biblically the rapture and the transformation of earth into hell.

### #trump 08/11:

keywords : goodbye, manchild, lost

that holds the lands of the blind!

of the brazen damned of public

happy retarded, nation, hmmmm

a little won, little lost

The poem generated phrases like “lands of the blind” and “of the brazen(,) damned of public”, phrases that could be used to describe what has happened during the Trump administration, such as the spread of misinformation (“lands of the blind”) and his shameless way of conduct (“of the brazen”) in the face of public backlash (“damned of public”). The last line “a little won, little lost” could mean the little gain the Trump administration has, and that there is little lost in the leaving of the administration.

### #trump 07/11:

Keywords: ouch, sad, harm

lost in a drab sorry world

a sad, guileless, sorry man

till he got sad and unhappy

they're sorry, they're sad

Looking at the poems, one can see that #trump 08/11 is more negative than #biden 08/11, most likely due to the much more negative significant words found in #trump tweets than #biden ones on the 8th. It is interesting to note how much more negative #trump 07/11 was compared to #trump 08/11, with all 4 lines denoting some sort of sadness. This may be due to the fact that tweets on the 7th were more primal in emotion, when Trump had just lost, and people are conveying pure emotion. This is contrasted with #trump 08/11, where the poem is less negative, but holds more meaning on the state of the country and the current situation. The use of the word “guileless” seem to indicate the sad man is not Trump, as one could not imagine him being devoid of guilt. As such, the narrative set by the poem of a sad man lost in a sad world seem to refer to Trump’s supporters, who were sold on the world Trump has convinced them of and how they are lost in it.

### #biden 08/11:

Keywords: finally, work, left, win

you can be the line in training

i vote, what you take of me

win to succeed them be

lying, robed in left remaining

Looking at #biden 08/11, it gives the sense of succession, while having a somewhat pyrrhic feeling with the last line “robed in left remaining”, as if saying that he has only received what was left remaining of his post. This implies that Trump has damaged the presidential role in some way, and Biden must now wear the rags that is the US president's image that Trump left behind. This could also imply the current situation of the United States, which was believed to be hit harder by the current pandemic due to mismanagement of the Trump administration. The lines “win to succeed them be lying” seems to be one phrase, with the word lying again implying dishonesty and how Biden is succeeding the liars, or Trump.

### #biden 07/11:

Keywords: lead, blunt, delivered

i will founder what to give

the free result of thy outcome

persons are most often players

had my share of site and place

As with #biden 08/11, #biden 07/11 also gives a feeling of succession. The first line “i will founder what to give” speaks of building up something or in this case maybe rebuilding what has been destroyed, as Biden founds his new administration over Trump’s, changing the previous administration policies and giving the people what he believes in. The second line seems to reference the right of free vote, implying that the outcome of the election was due to the people’s free will in voting. This comments on the claims of voting fraud Trump and his party are insisting on till this day.

## Discussion:

The use of big data techniques were met with several problems. The use of a large corpus of text required specialized code and function that were not readily available. The maintenance of system ram also was a concern due to the data structures used. Generators and batch processing techniques were used to mitigate these problems. The use of more advanced techniques like word embeddings and regression models could have also aided in producing more varied results.

Further analysis into different methods of poetry generation may net better results due to our current use of PoeTryMe as a “Black Box”, further research and development would be necessary for the creation of new poetry generation techniques.

The poetry generated from the corpus was typically irrelevant requiring many passes through the generator to produce one of good enough quality and relevance. The methods for generating the poetry were limited to manually inputting the seed words hence the manual curation of the poems was done at generation. Selecting “good” poetry was a humanistic task and would probably be difficult to automate.

Ultimately the task of analysing the poetry is left to the human. New texts can be generated through machines but in order to make sense and draw conclusions from these new texts, an expert human reader and analyser is still very much required.

## Conclusion:

While there were many challenges in producing meaningful results from this method, from both the processing of the big data to the limitations of the poetry generator, this was an interesting experiment in analysing data not through conventional digital means of numbers and statistics but by converting it to a more humanistic form. While Digital humanities has been using digital methods to create metadata from humanistic works to help study the work, this paper shows that there is also room for the inverse; for big data to be converted to a more humanistic form and be used as a supplement to the digital analysis of big data.

1. https://monkeylearn.com/blog/what-is-tf-idf/#:~:text=TF%2DIDF%20is%20a%20statistical,across%20a%20set%20of%20documents. [↑](#footnote-ref-1)
2. https://www.kaggle.com/manchunhui/us-election-2020-tweets [↑](#footnote-ref-2)
3. <https://poetryme.dei.uc.pt/> [↑](#footnote-ref-3)
4. https://monkeylearn.com/blog/what-is-tf-idf/#:~:text=TF%2DIDF%20is%20a%20statistical,across%20a%20set%20of%20documents. [↑](#footnote-ref-4)
5. Hugo Gonçalo Oliveira. PoeTryMe: a versatile platform for poetry generation. In Proceedings of the ECAI 2012 Workshop on Computational Creativity, Concept Invention, and General Intelligence, C3GI 2012, Montpellier, France, August 2012. [↑](#footnote-ref-5)