# Comparison Study of Misinformation in Authoritarian and Non-Authoritarian Countries

## DSKUS Team 3 Final Report

Andrea Chung

Hyunji Jeong

William McWhorter

Yesung Shin

June 13, 2023

## 1. Introduction

Online social media's recent growth has made it easier for users to share information not only accurate ones but also prejudiced ones. Following the COVID-19 epidemic, fraudulent news about the disease was wildly disseminated through social network sites (SNS), making it more difficult to tell the difference between the true and false reports. As people come across false information about COVID-19, the accurate information becomes confounded, which would eventually give negative impact on society as well as the pandemic.

The inevitability and importance of confrontation behind the issue of fake news has led to active and fruitful studies until today. However, there are more to be evaluated and improved. Various interpretations and definitions are given on defining fake news (Chu, Xie, & Wang, 2021) but for this study, the following description from Chu et al. is used: "fabricated news that is written to mislead readers or attract people's attention (Chu, Xie, & Wang, 2021)." Especially during the time of COVID-19, the terms like 'infodemic' arise due to dissemination of information that is not only factual but also feigned (Pang, Liu, & Lu, 2022). The issue of misinformation spread on social media is on global scale as people can easily be connected to each other (Pang, Liu, & Lu, 2022).

For the project, comparative studies of misinformation in authoritarian and non-authoritarian countries were done. China and United States of America (USA) is chosen to be the representative of totalitarian and non-totalitarian countries and its iconic social media platforms were chosen for each country. Twitter and Weibo were selected as they are the most used ones in each nation. To conduct the study, the following hypothesis is given:

*The authoritative countries such as China tend to be more controlling on social media platforms regarding fake news than non-authoritative countries such as* *the U.S. The two divergent characteristics can be reflected through two social media platforms—Twitter and Weibo.*

Two primary research questions are proposed as the following:

1. What are the similarities and differences in the characteristics of fake news between two countries?
2. How does social media from two countries play differently in spreading the fake news?

To answer the proposed research question, various in-depth comparison analyses are done.

## 2. Literature Review

Detecting false news has been the subject of numerous studies and research projects in recent years. The ways in which people can consume information have increase with popularity of social media. We may come across a variety of information, including accurate yet biased information. There have been many studies on detection models of fake news and Reis et al. (Reis, Correia, Murai, Veloso, & Benevenuto, 2019) compares several supervised learning algorithms with several features to create a model to distinguish the fake news. The model was trained using linguistic features, environmental characteristics, and news source features. Additionally, comparison studies and its application on cross-language detection model has been done. Chu et al. (Chu, Xie, & Wang, 2021) explores on the conveyance of fake news recognition model across different languages such as Chinese and English.

Furthermore, the study conducted by Pang et al. (Pang, Liu, & Lu, 2022) also reveals a thorough examination of Chinese social media sites Weibo and WeChat. Both were discussed in terms of “socially mediated public spheres” and provided insights into how information is communicated in totalitarian nations like China. Even though there have already been many studies on subject, more need to be done in the future. Particularly, additional research on social media platforms in authoritarian nations and comparison studies with non-authoritarian nations like USA are necessary.

## 3. Exploratory Data Analysis

### *3.1. Twitter*

For the exploration of fake news on Twitter, labelled datasets from two different published papers are combined and utilized. The first dataset comes from Shahi et al (Shahi, Dirkson, & Majchrzak, 2021). and the second one comes from Das et al (Das, Basak, & Dutta, 2021). The labels are adjusted to agree and merge without difficulties. As part of the preprocessing stage, all the missing rows are removed. Using 'snscrape' package, auxiliary information such as number of comments, number of reposts, and more on tweets are collected. The texts of the tweets are processed and cleaned by eliminating punctuation, stop words, URLs, emoticons, and other unnecessary linguistic barriers before the deeper analysis on the texts. The distribution of the "True" and "False" labels is shown in the Figure 1.

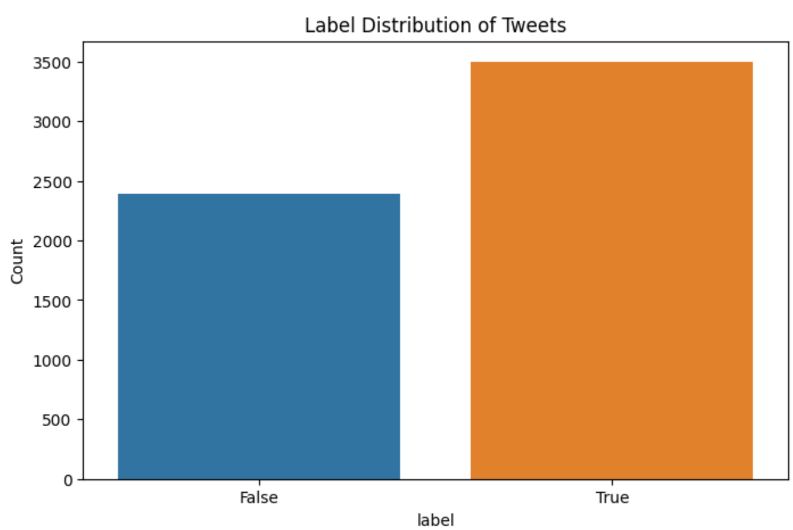



Figure 1. Label Distribution of Twitter Dataset

Initial analysis considers 18 attributes, including replyCount, retweetCount, likeCount, quoteCount, source, number of links, number of used media, location, userVerified, userFollowerCount, userFollowingCount, userTweetCount, userFavoritesCount, userListedCount, userMediaCount, word count, polarity, and accountAgeAtPosting (days). The interquartile range (IQR) approach is used to eliminate outliers from the data. With the data divided between true and fake labels, each of the attributes are visualized with either violin plot or bar plot.

Among the selected attributes, only a few of them exhibited significant differences between the labels. Still, small differences can be noticed for each feature. In general, tweets with label “True” were slightly more likely to have higher values of replyCount, rewteetCount, likeCount, quoteCount, userFollowerCount, and polarity. Additionally, on average, the true labelled tweets had longer content or higher word counts, which can be depicted from the Figure 2.

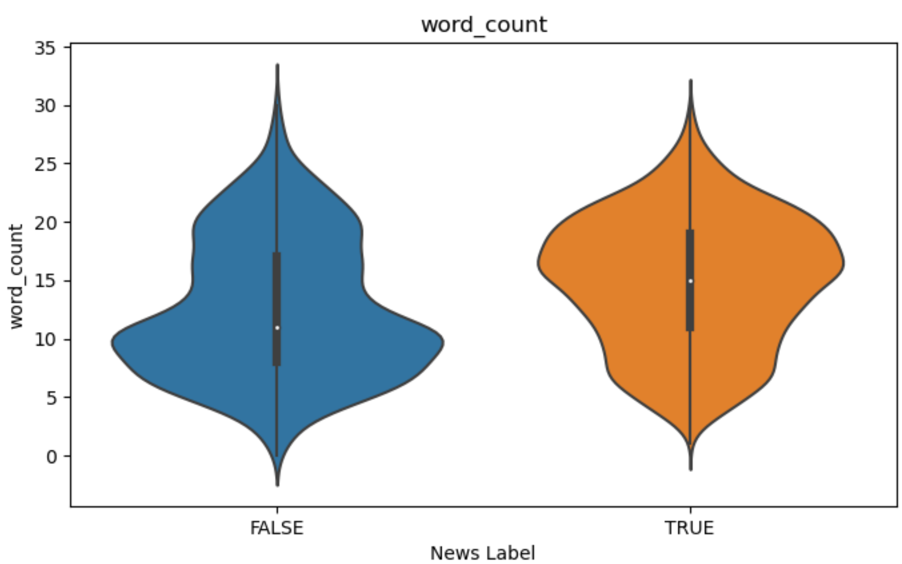



Figure 2. Word Count by Tweet Labels (True, False)

One of the interesting findings is that the false labeled tweets are inclined to be from unknown sources than tweets with true labels when satirical sites are ignored. Furthermore, there is strong association between user's verification status and the true status of a tweet: verified users are almost exclusively classified as true. This finding is shown in Figure 3.

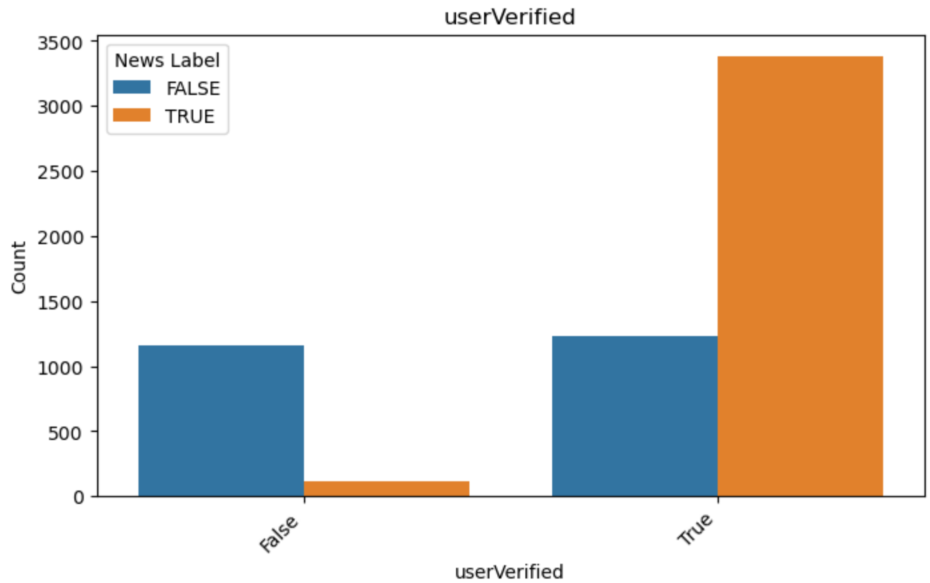



Figure 3. Distribution of User Verification in Twitter Data

Furthermore, the parts of speech (POS) are extracted from the tweets and had their correlations mapped with the label of the tweets. In general, the label of the tweet did not directly correlate with frequency of certain POS was used. Plural nouns, present participle verbs, and past participle verbs, on the other hand, have minute relationship with the labels. The correlation heatmap can be seen in Figure 4.

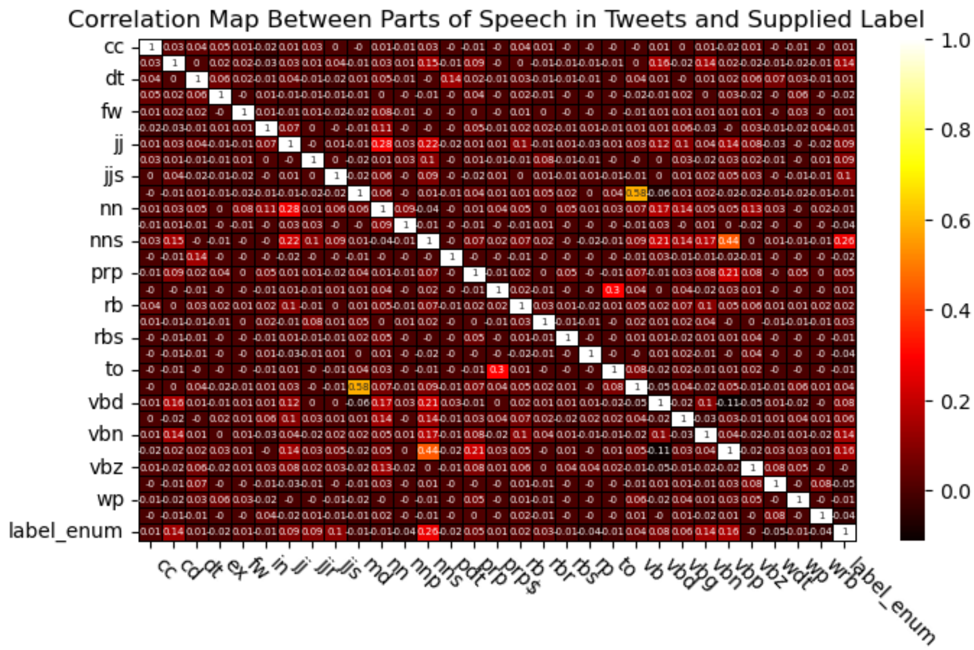



Figure 4. Correlation Map Between POS in Tweets and Labels

Finally, the distribution of words within the tweets was analyzed. The word clouds for each label, true and false, can be seen in Appendix (A) and Appendix (B). From the analysis, false tweets can be seen to utilize terms like “lockdown” and “claim” frequently, and uniquely use terms such as “#donaldtrump”, “lie”, “biden”, and “fake”. The two tables, Table 1(a) and 1(b), show the most distinctive words for each label or words that primarily show up in either true or false labels. Interestingly, the linguistic uniqueness shows that tweets with assertive tone or with political undertone are more likely to be classified as misinformation.

Top 20 terms that appear primarily only in TRUE labels

| | word_in_tweet | frequency_ratio |
|---|---|---|
| 56 | auckland | 76.164556 |
| 81 | tracked | 57.286162 |
| 83 | #indiawillwin | 56.635183 |
| 108 | states | 46.219517 |
| 115 | 7day | 44.917559 |
| 141 | england | 37.105809 |
| 173 | #covidview | 31.246997 |
| 169 | statesuts | 31.246997 |
| 171 | percentage | 31.246997 |
| 200 | completed | 28.643081 |
| 216 | #kayburley | 26.039164 |
| 226 | steady | 25.388185 |
| 228 | improves | 24.737206 |
| 229 | returned | 24.737206 |
| 58 | managed | 24.086227 |
| 42 | hospitalization | 23.923482 |
| 246 | miq | 23.435248 |
| 245 | improvement | 23.435248 |
| 34 | active | 23.174856 |
| 251 | fct | 22.784269 |

(a) Table of terms that are primarily used in tweets labeled as true and not in tweets labeled as false.

Top 20 terms that appear primarily only in FALSE labels

| | word_in_tweet | frequency_ratio |
|---|---|---|
| 5839 | #donaldtrump | 0.007154 |
| 5840 | #coronavirusfacts | 0.011625 |
| 5841 | misinformation | 0.016692 |
| 5842 | biden | 0.017594 |
| 5843 | #factcheck | 0.017594 |
| 5844 | #fakenews | 0.018083 |
| 4925 | fake | 0.018599 |
| 5845 | #coronaoutbreak | 0.028303 |
| 5846 | truth | 0.029590 |
| 4969 | kill | 0.031755 |
| 5847 | #health | 0.032549 |
| 1127 | claim | 0.032961 |
| 4271 | #china | 0.033384 |
| 5850 | rumor | 0.034262 |
| 5848 | blame | 0.034262 |
| 5849 | photo | 0.034262 |
| 1910 | corona | 0.036166 |
| 5851 | conspiracy | 0.036166 |
| 1092 | president | 0.036935 |
| 5852 | #coronacheck | 0.038293 |

(b) Table of terms that are primarily used in tweets labeled as false and not in tweets labeled as true.

Table 1

### *3.2. Weibo*

For Chinese dataset, CHECKED dataset by Yang et al (Yang, Zhou, & Zafarani, 2021). is utilized. It is the first Chinese dataset on COVID-19 misinformation. The dataset had 1776 Weibo posts with "real" label and 344 Weibo posts with "fake" label. The columns of the raw data are the following: label, id, date, time, user_id, text, pic_url, video_url, comment_num, repost_num, like_num, analysis. The columns of pic_url and video_url was encoded as categorical variable, which "1" represents the presence of URLs and "0" showing the opposite. To solve the issue of data imbalance between real and fake labels, random sampling technique is used on posts with real label.

Looking at the correlation between features, one noticeable finding is that fake news, unlike real news, has considerably higher correlation with like_num (number of likes) and repost_num (number of repost). Also, relatively high association with repost_num and comment_num can be seen. The correlation matrix is shown in Figure 5. The finding suggests that spreading misinformation on social media can be done by raising number of reposts and likes, which strengthens the dispersal phenomena.

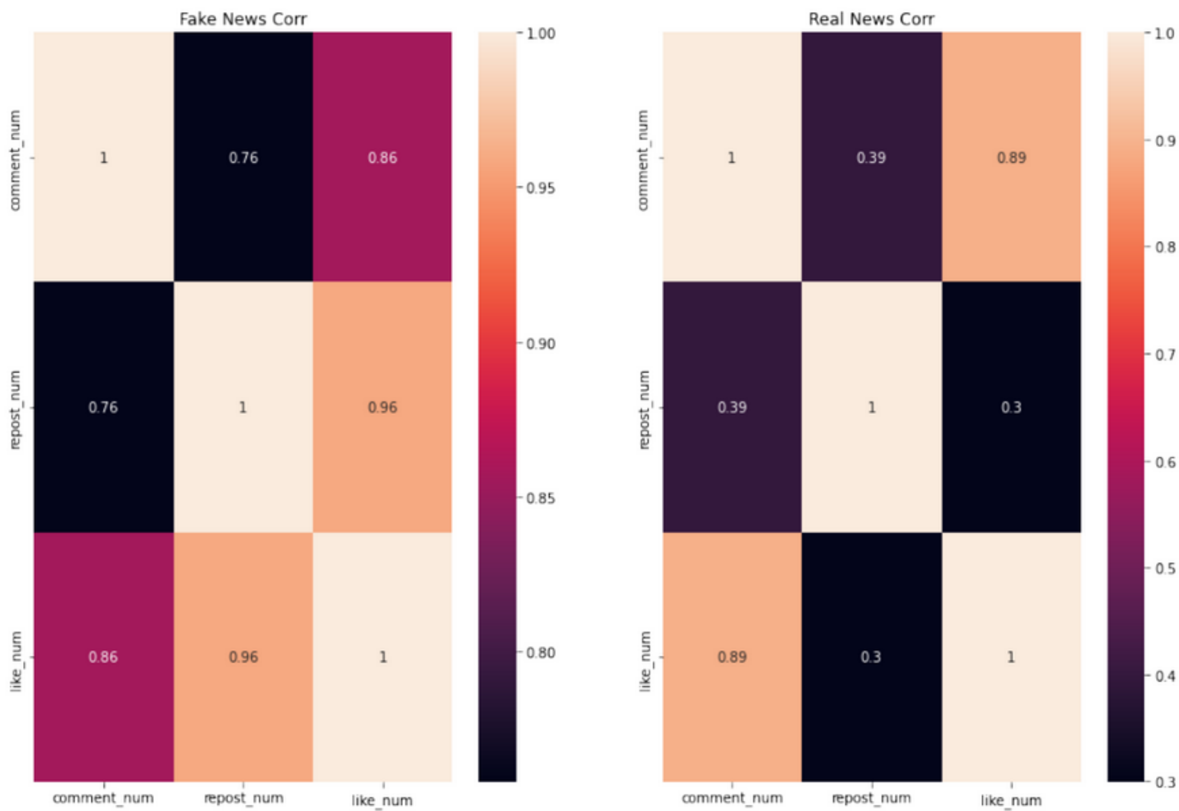



Figure 5. Correaltion Matrix of Numeric Features of Weibo Dataset

Like Twitter dataset, the outliers are removed using IQR method and the natural language processing (NLP) techniques are applied to clean the given text. Unlike twitter tweets, Weibo posts are originally written in Chinese, which demands for the need of translation before in-depth analysis. Thus, DeepL translation API was used after the comparison with google translator as it manifested higher accuracy in translation. The preprocessing techniques are applied to the translated texts.

Additionally, statistical linguistic measures are visualized and can be seen on Figure 5 and 6.

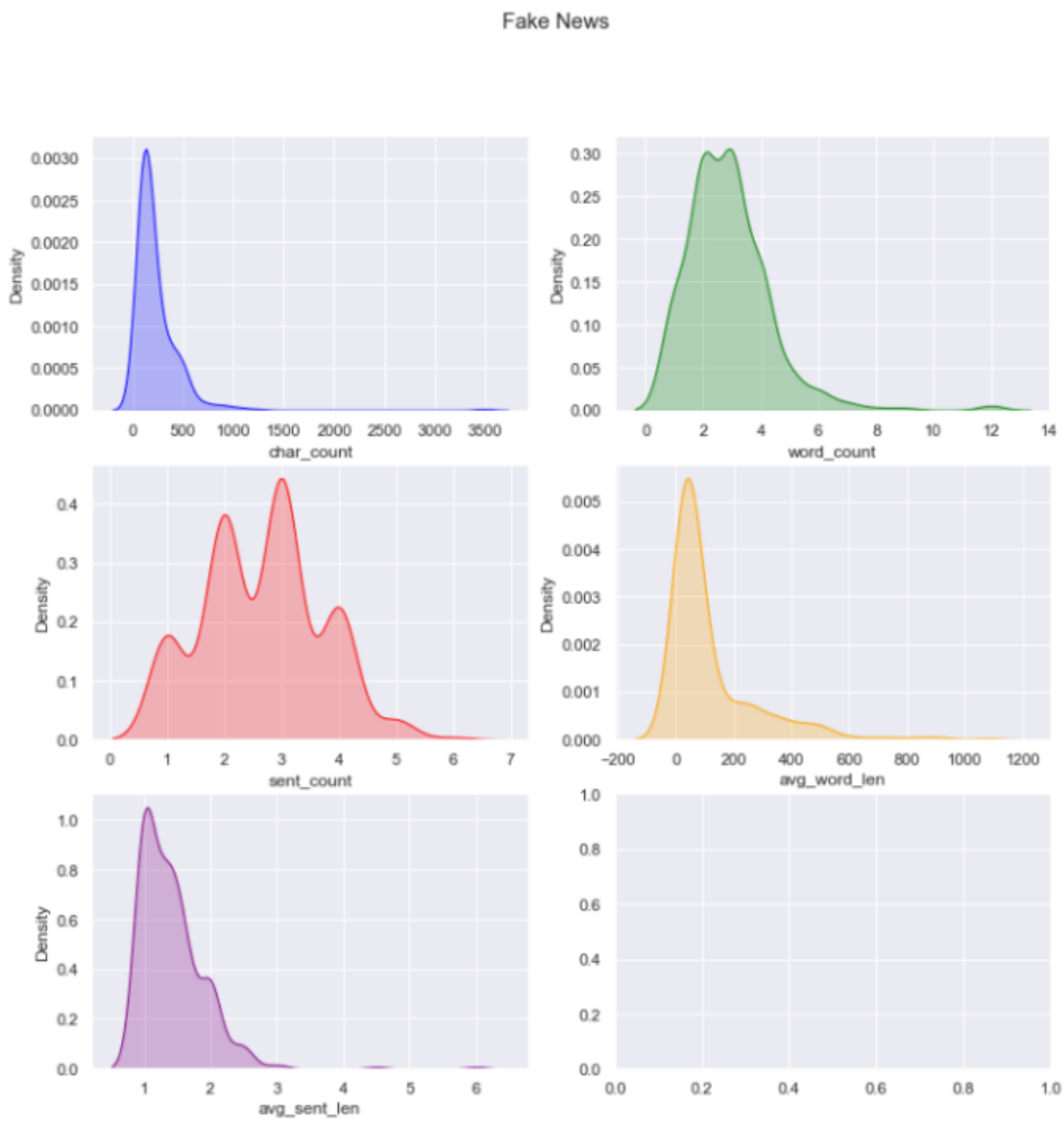



Figure 6. Numeric Analysis of Linguistic Features on Weibo Fake News Data

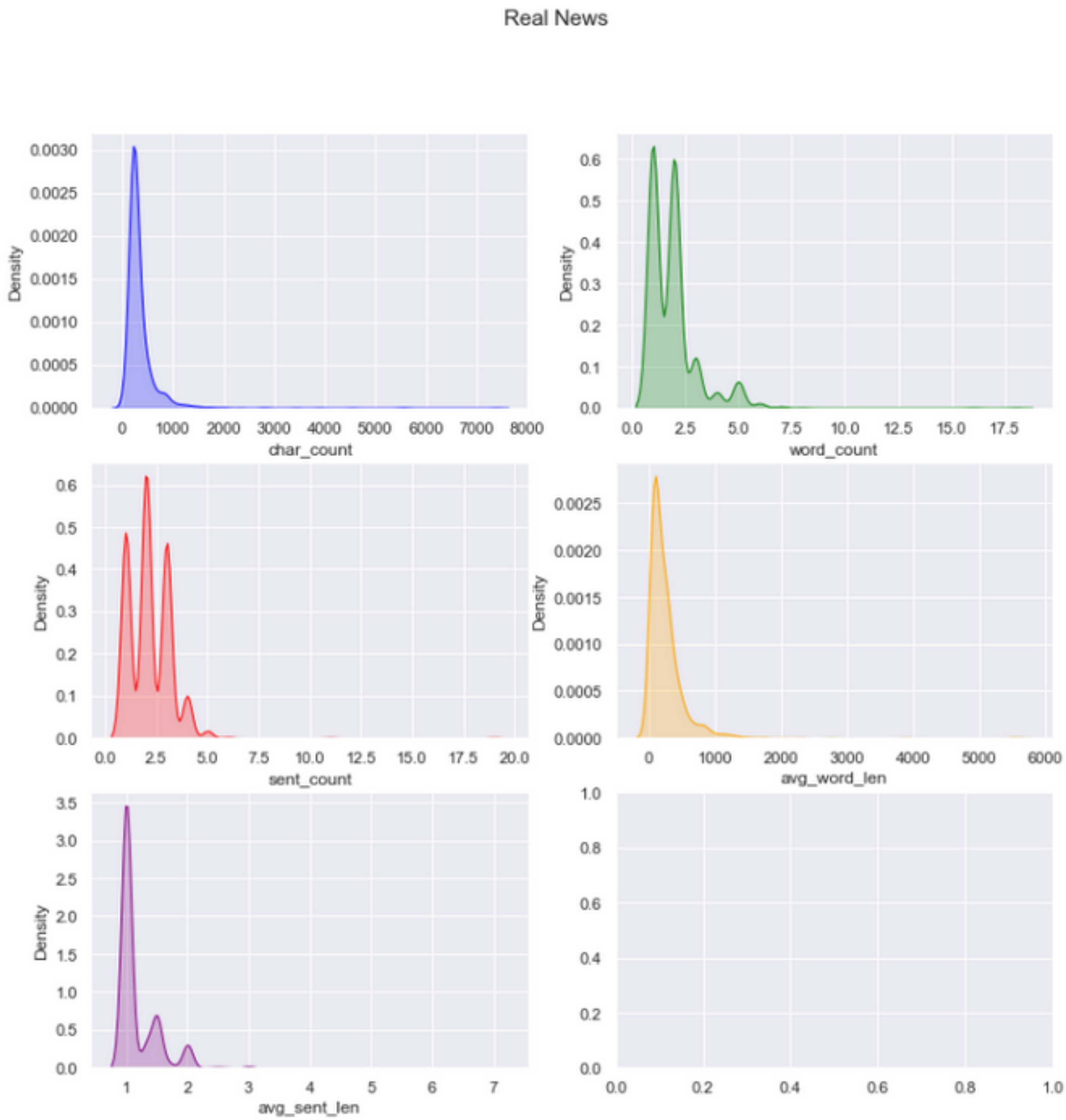



Figure 7. Numeric Analysis of Linguistic Features on Weibo Real News Data

The distribution of words within the Weibo posts was analyzed by looking at uniquely used terms for each label. The two tables, Table 2(a) and 2(b), show the most particularly used terms.

Top 20 terms that appear primarily only in TRUE labels

| | word_in_weibo | frequency_ratio |
|---|---|---|
| 51 | coronary | 67.970653 |
| 34 | observation | 42.981442 |
| 95 | provincial | 38.892299 |
| 123 | centralized | 29.805313 |
| 125 | data | 28.169656 |
| 126 | cdc | 27.624436 |
| 127 | autonomous | 27.442697 |
| 134 | relevant | 26.715738 |
| 138 | urumqi | 26.170519 |
| 77 | dalian | 24.898341 |
| 78 | cured | 23.262683 |
| 161 | john | 23.080944 |
| 163 | hopkins | 22.717464 |
| 173 | present | 21.808766 |
| 176 | jilin | 21.627026 |
| 189 | august | 20.173108 |
| 88 | reporter | 20.173108 |
| 0 | case | 19.014089 |
| 61 | cumulative | 18.900930 |
| 99 | symptom | 18.628321 |

(a) Table of terms that are frequently used in Weibo posts labeled as True

Top 20 terms that appear primarily only in FALSE labels

| | word_in_weibo | frequency_ratio |
|---|---|---|
| 8887 | xinjia | 0.004783 |
| 8888 | beichuan | 0.004912 |
| 8889 | sailed | 0.006491 |
| 8890 | dog | 0.007270 |
| 8891 | sealed | 0.009087 |
| 8892 | hunt | 0.009565 |
| 8893 | coal | 0.009565 |
| 6363 | tangshan | 0.010385 |
| 8894 | tiger | 0.010691 |
| 8895 | ancestor | 0.010691 |
| 4127 | normalization | 0.011015 |
| 8897 | xiaoyue | 0.011359 |
| 8898 | mosquito | 0.011359 |
| 8896 | ah | 0.011359 |
| 8900 | threeyear | 0.012116 |
| 8899 | zhengli | 0.012116 |
| 8901 | starvation | 0.012116 |
| 8436 | harbor | 0.013980 |
| 5070 | warship | 0.015145 |
| 8903 | remdesivir | 0.016522 |

(b) Table of terms that are frequently used in Weibo posts labeled as False

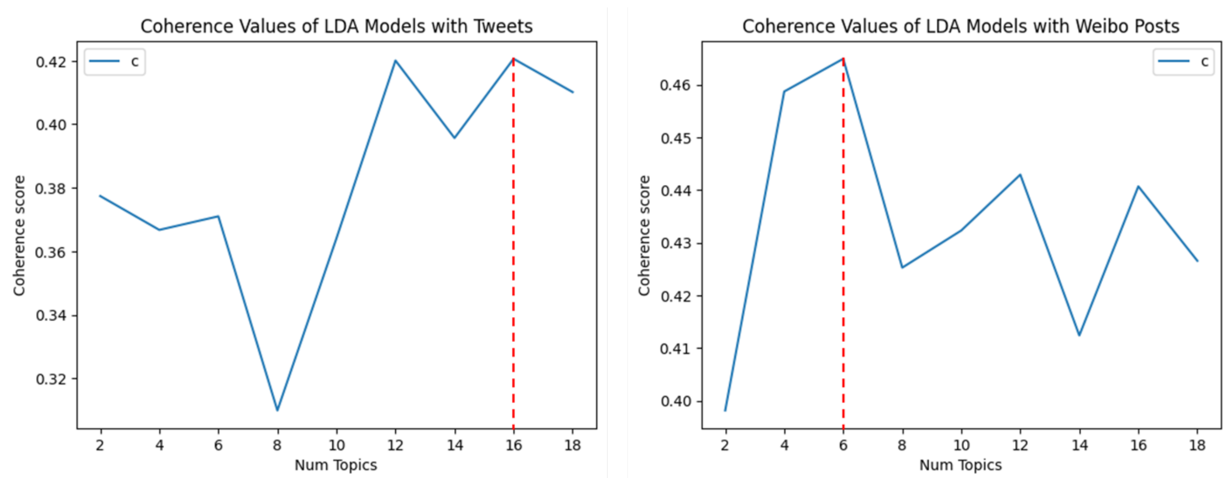
Table 2

One of the interesting findings is that frequently appearing terms in fake posts tend to be weakly related to the topic of COVID-19 compared to frequently appearing words in real posts. The terms such as “cdc”, “coronary”, and “symptom” are highly related to the epidemic but terms like “dog”, “tiger”, and “ancestor” comparatively has lower relatability to COVID-19. Unlike Twitter dataset, Weibo, from the initial EDA, tends to have “factual” or “technical” terms when showing accurate information. In addition, the separate word clouds for each labels can be found in Appendix (C) and (D).

## 4. Analysis

### *4.1. Topic Modelling*

As part of initial comparison analysis of text classification, topic modelling is done using Latent Dirichlet Allocation (LDA) for both Twitter and Weibo datasets. The coherence score is used to determine the optimal number of topics to be considered when employing topic modelling analysis. Hyperparameter tuning is done to show appropriate number of topics to be determined and notable finding is that prime number of topics for Weibo posts are comparatively much lower than Twitter tweets. The finding can be visualized with the Figure 8 below:




(a). Coherence Values of LDA Models with Tweets

(b) Coherence Values of LDA Models with Weibo Posts

Figure 8

The ideal number of topics for Tweets is 16 when the Weibo is 6. The finding suggests that types and number of subjects discussed in the Twitter tends to be more diverse than Weibo. In authoritarian country like China, the social media can be referred as “socially mediated public space” as Pang et al. indicates. The diversity of topics is very limited in Weibo compared to Twitter, which displays divergent characteristics of two social media platforms.

*4.2. Text Classification*

To implement text classification, a few various machine learning classifier algorithms are used with vectorized text with term frequency-inverse document frequency (Tf-IDF). The accuracy score of Logistic regression (LR), multinomial naïve bayes (MNB), random forest classifier (RFC), and gradient boosting classifier (GBC) are compared to find the most suitable model for text classification. When hyperparameter tuning was done and the models were applied to both Twitter and Weibo datasets, the model with least difference in accuracy score (Table 3) between the two data was RFC. However, the accuracy except for GBC model on Weibo data had not much difference relatively.

| | Twitter | Weibo |
|---|---|---|
| Logistic Regression | 88.62 | 90.57 |
| Multinomial Naive Bayes | 90.99 | 86.75 |
| Random Forest Classifier | 89.97 | 91.42 |
| Gradient Boosting Classifier | 88.70 | 51.23 |

Table 3. Classifier Model Comparison using Accuracy Score

Furthermore, the comparison between textual feature importance is done by deploying feature importance from RFC model. The important features show the most significant tokens or terms that highly supports the performance of the classifier model. The following two tables, Table 4(a) and 4(b), display the top 10 important textual features for each social media. The important features does not exhibit much difference between the two social media platforms, which shows the possibility of employing same classifier model for both platforms.

```
Top 10 Important Features:
case: 0.013549401750247672
covid: 0.012872345737428374
state: 0.011192918858542915
coronavirus: 0.00917016367700809
trump: 0.008053006661541755
number: 0.007440993980845244
test: 0.007010737533716628
news: 0.006764544063522971
testing: 0.006567840865060019
data: 0.006154610149442493
```

(a) Top 10 Important Textual Features from Twitter Tweets

```
Top 10 Important Features:
case: 0.013769063452835874
covid: 0.011638255367693235
state: 0.010931993205839733
coronavirus: 0.009841576957607264
news: 0.0078930709417204
trump: 0.007529707716651101
test: 0.00724424436392183
number: 0.007241696427199621
indiafightscorona: 0.00680208467086783
testing: 0.006592510934797018
```

(b) Top 10 Important Textual Features from Weibo Posts

Table 4

*4.3. Sentiment Classification*

For sentiment classification, the column, ‘sentiment’, is created with 'positive' and 'negative' label to divide real and fake news. For Twitter, there were 3496 rows of 'positive' label which is real news, and 2388 rows of 'negative' label which is fake news (Figure 9). For Weibo, there were 1760 'positive' labels and 344 'negative' labels (Figure 10).

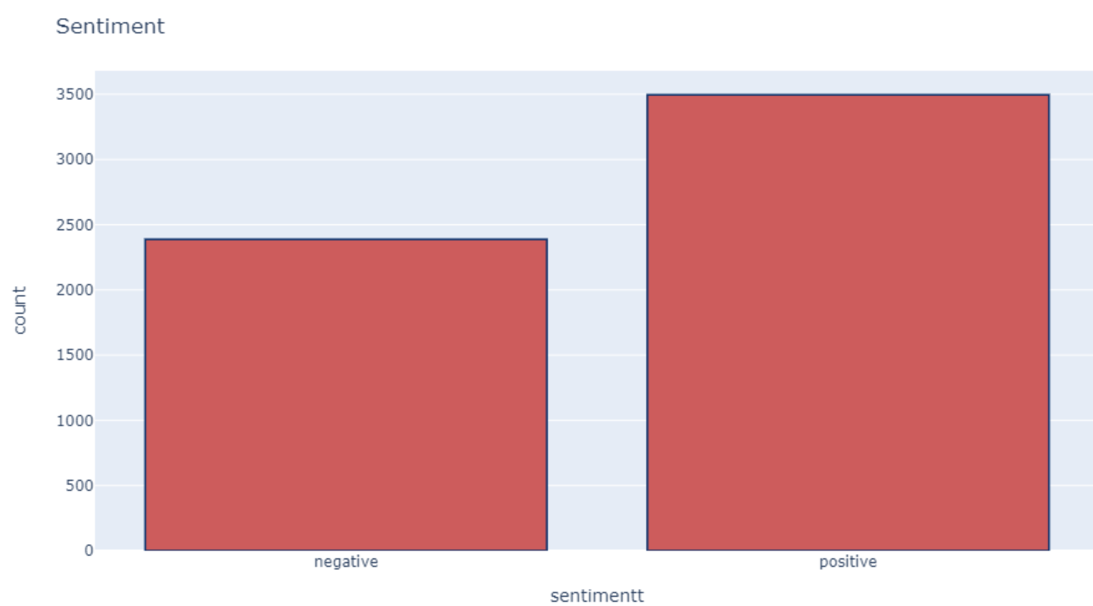



Figure 9. Twitter Sentiment Plot

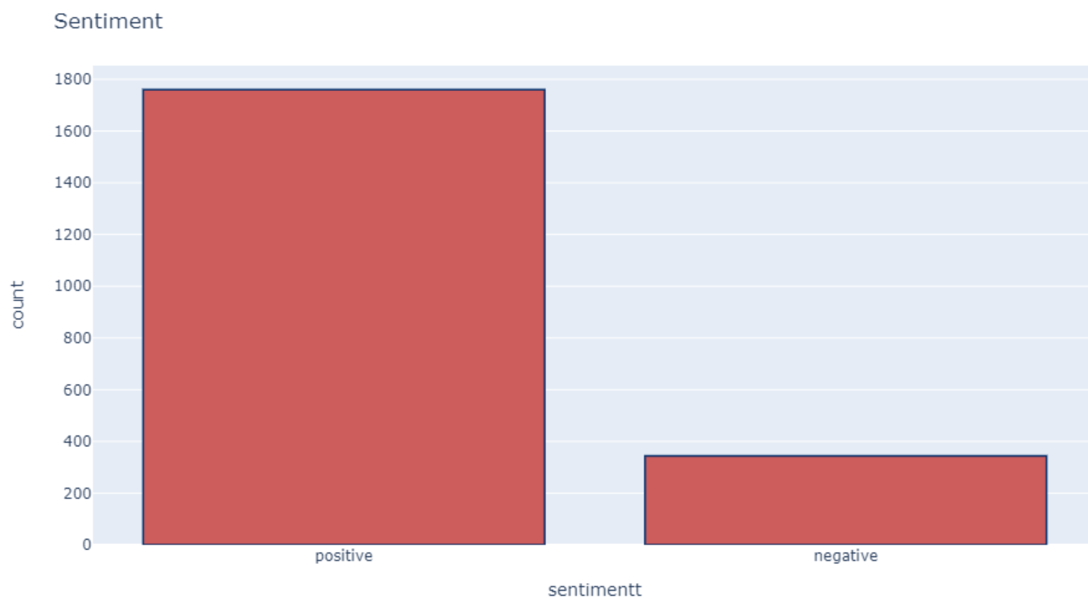
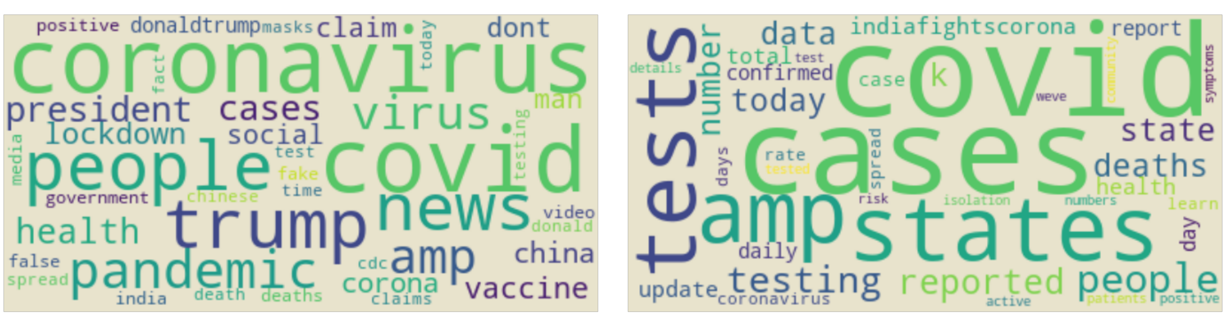



Figure 10. Weibo Sentiment Plot

Further comparison analysis was done using Word Cloud. The noticeable finding is that Weibo posts displays apparently observable differences between fake and real news. However, Twitter tweets, in general, does not show big different between the sentiments. The difference is visualized through Word Cloud (Figure 11(a) and (b)) and the terms ‘China’ and ‘lockdown’ appears a lot, which can be considered as biased words.




(a) Twitter Fake News Word Cloud (b) Twitter Real News Word Cloud

Figure 11

### *4.4. Feature Importance Analysis*

As part of comparative analysis, feature importance analysis is performed by two different ways. Both coefficient of tree-based model and permutation importance are obtained and analyzed. First, looking at the feature importance coefficients acquired from Random Forest Regressor (RFR) model, the following three attributes played the biggest role for Twitter dataset (Figure 12): “likeCount”, “month”, and “char_count” (number of characters).

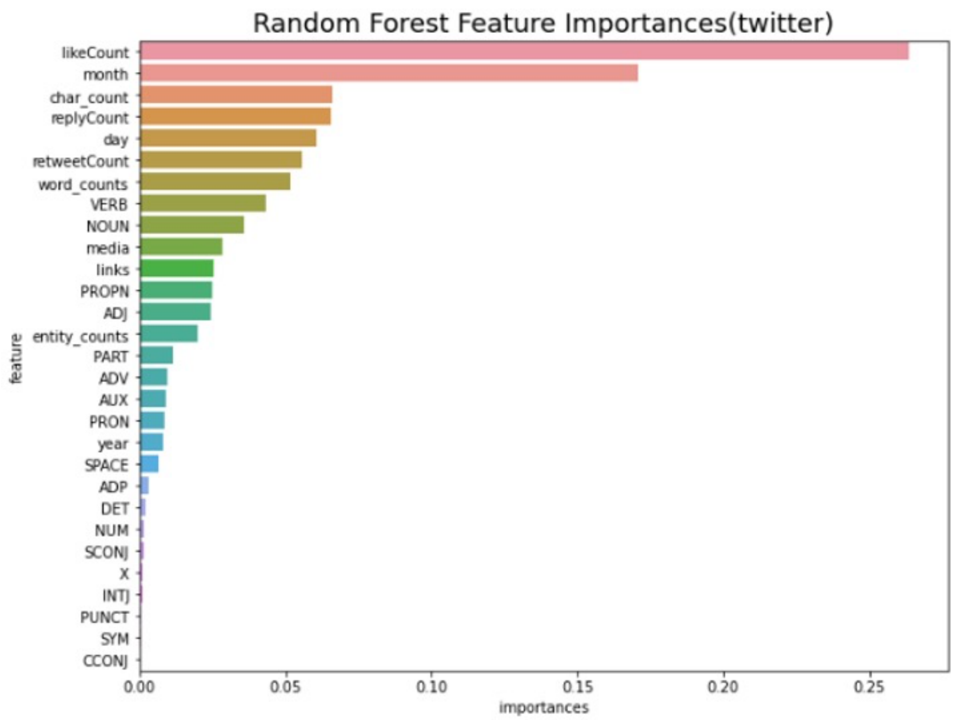



Figure 12. Coefficient-based feature importance of RFR model for Twitter Dataset (w/out "tweetID")

The feature “month” is one of the key attributes, which shows the inevitable impact of current events in our society, which the issues vary by month to month. Also, the number of likes plays a significant role and displays the inherent characteristic of social media platform, where high probability of dispersion of the news results from high number of likes. Likewise, similar important features can be observed for Weibo. The top three attributes are: “like_num”, “repost_num”, and “comment_num” (Figure 13). The similarity between two social media platforms show that characteristics of social media platform does not vary much despite its difference in country and its government system.

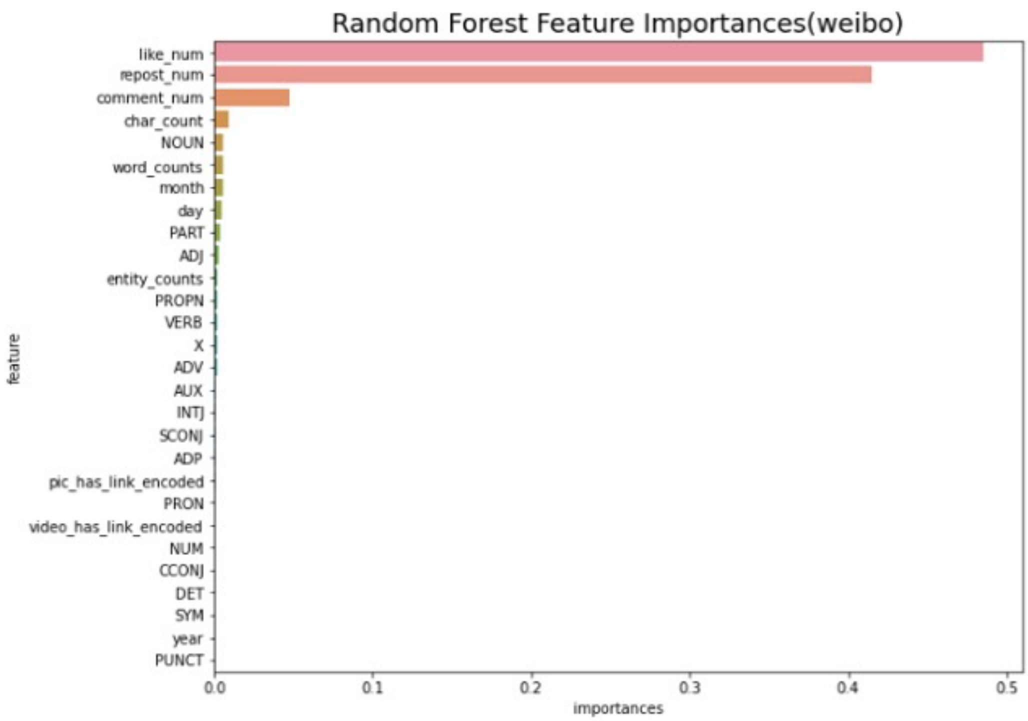



Figure 13. Coefficient-based feature importance of RFR model for Weibo Dataset

Another feature importance analysis is done using permutation feature importance technique. The overall result does not show significant difference from RFR feature importance method. However, the weights for important attributes are relatively much more considerable for Weibo dataset than Twitter dataset (Table 5). The weight of "month" quality from Twitter data is 0.4135 when the weight of "like_num" aspect is 0.7591. The finding suggests unknown or unexplicit attributes that may be consequential for Twitter. In general, the many similar aspects are shared with two social media platforms and number of comments and reposts are top features to be considered.

| Weight | Feature |
|---|---|
| 0.4135 ± 0.0534 | month |
| 0.3760 ± 0.0291 | likeCount |
| 0.1941 ± 0.0220 | replyCount |
| 0.0903 ± 0.0106 | links |
| 0.0873 ± 0.0047 | day |
| 0.0645 ± 0.0022 | retweetCount |
| 0.0642 ± 0.0103 | char_count |
| 0.0459 ± 0.0086 | VERB |
| 0.0416 ± 0.0063 | media |
| 0.0397 ± 0.0023 | word_counts |
| 0.0305 ± 0.0072 | NOUN |
| 0.0188 ± 0.0046 | PROPN |
| 0.0129 ± 0.0014 | PART |
| 0.0126 ± 0.0041 | ADJ |
| 0.0106 ± 0.0026 | entity_counts |
| 0.0102 ± 0.0056 | year |
| 0.0074 ± 0.0025 | AUX |
| 0.0054 ± 0.0009 | PRON |
| 0.0043 ± 0.0012 | ADV |
| 0.0024 ± 0.0016 | NUM |
| … 9 more … | |

(a) Permutation-based feature importance of RFR model for Twitter

| Weight | Feature |
|---|---|
| 0.7591 ± 0.0790 | like_num |
| 0.6964 ± 0.0451 | repost_num |
| 0.3847 ± 0.0379 | comment_num |
| 0.0064 ± 0.0021 | PART |
| 0.0054 ± 0.0093 | char_count |
| 0.0039 ± 0.0041 | NOUN |
| 0.0030 ± 0.0031 | month |
| 0.0012 ± 0.0008 | entity_counts |
| 0.0011 ± 0.0005 | pic_has_link_encoded |
| 0.0008 ± 0.0004 | ADV |
| 0.0007 ± 0.0006 | ADJ |
| 0.0007 ± 0.0007 | day |
| 0.0006 ± 0.0023 | X |
| 0.0003 ± 0.0016 | INTJ |
| 0.0002 ± 0.0008 | SCONJ |
| 0.0002 ± 0.0003 | PRON |
| 0.0000 ± 0.0029 | word_counts |
| 0 ± 0.0000 | NUM |
| 0 ± 0.0000 | year |
| 0 ± 0.0000 | PUNCT |
| … 8 more … | |

(b) Permutation-based feature importance of RFR model for Weibo

Table 5

*4.5. Clustering Analysis*

Lastly, comparative analysis using clustering techniques are done. The Weibo dataset is reduced to only include the features representing the number of comments, number of reposts, the number of likes, and the lemmatized post text. The Twitter dataset is similarly reduced to features representing the number of replies, number of retweets, number of likes, number of links, number of media attachments, and the lemmatized tweet text. The lemmatized texts are converted into a TF-IDF vectorization format, and the sentiment polarity of the posts is calculated. Also, the statistical columns are standardized to make features to be on the same scale.

Clustering is performed using both k-means and hierarchical clustering techniques. The hierarchical clustering is done by agglomerative clustering with ward linkage, complete linkage, average linkage, and single linkage. These five clustering techniques are performed over both datasets, and both datasets are split into two other datasets: one including only the TF-IDF vectorization of the posts and one adding the post statistics to the TF-IDF vectorization.

K-means is performed for all cluster sizes from 1 through 20, and the sum of squared errors (SSE) is recorded for Twitter and Weibo, both with and without post statistics. As evidenced by the SSE curve in Figure 14 and the utilization of the elbow method to determine appropriate number of clusters, the datasets without post statistics yield lower SSE values and have no preference for the number of clusters. As a result, it was chosen to utilize two clusters moving forward for k-means and for hierarchical clustering so that the results could be compared to the true and false news labels of the datasets.

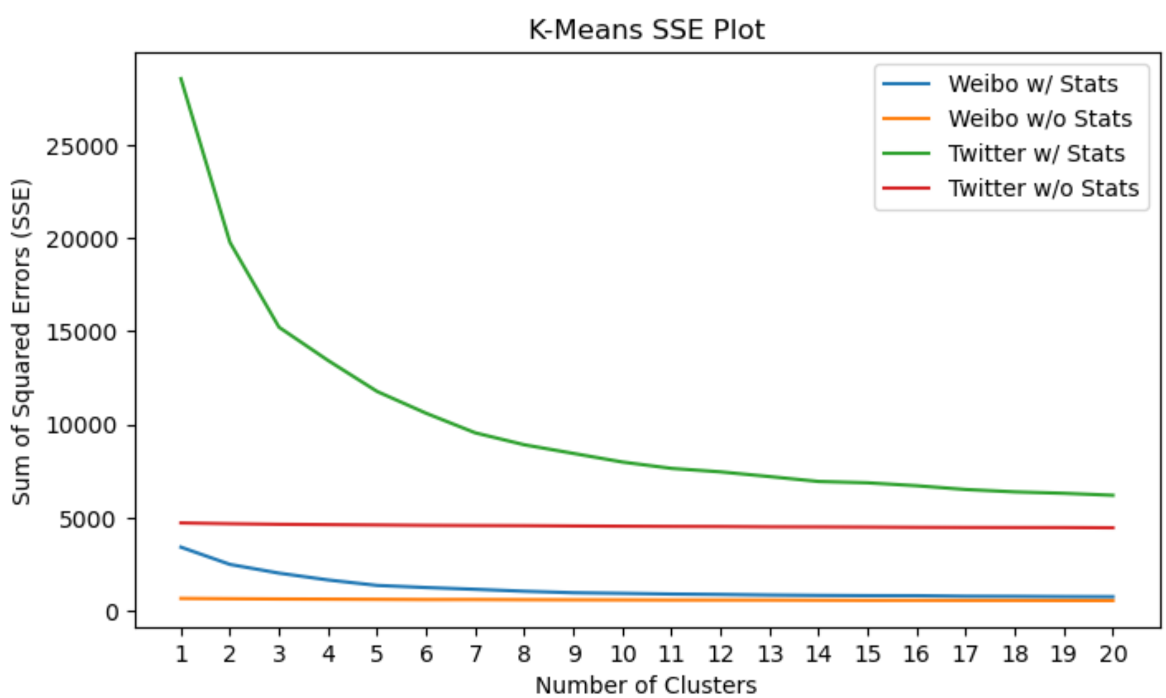



Figure 14. SSE Curves for Twitter and Weibo Datasets using K-Means

The results of the clustering were compared to the true and false labels of the datasets. Since clustering was chosen to result in two clusters, combining clusters for the comparison was not necessary. The clusters were assigned such that the highest accuracy was yielded, as otherwise it could be interpreted as the labels being placed backwards. With the clusters and labels determined, the accuracy, precision, recall, and F1 score of the clustering was calculated. These results are shown in Tables (6) through (9). The accuracies for both datasets and both feature sets are generally poor, however the feature sets comprised of only the TF-IDF values tended to perform equal or better than the feature sets including post statistics. K-means tend to have better performance with regards to accuracy (Table 6), but agglomerative clustering using average linkage tend to perform better with regards to the F1 score (Table 9). To easily visualize the findings, four Word Clouds for Weibo and Twitter datasets without post statistics for both k-means and average linkage agglomerative clustering can be seen from Appendix (E) to (H).

| | weibo_stats | weibo_no_stats | twitter_stats | twitter_no_stats |
|---|---|---|---|---|
| k-means | 0.513081 | 0.626453 | 0.503565 | 0.611158 |
| ward | 0.504360 | 0.597384 | 0.501678 | 0.512164 |
| complete | 0.501453 | 0.578488 | 0.500419 | 0.503356 |
| average | 0.501453 | 0.500000 | 0.501678 | 0.500210 |
| single | 0.501453 | 0.501453 | 0.500210 | 0.500210 |

Table 6. Accuracies of 5 Clustering Techniques Comparing to True and False Labels

| | weibo_stats | weibo_no_stats | twitter_stats | twitter_no_stats |
|---|---|---|---|---|
| k-means | 1.000000 | 0.957895 | 0.498209 | 0.920635 |
| ward | 0.000000 | 0.958904 | 0.499159 | 1.000000 |
| complete | 1.000000 | 0.442060 | 0.500210 | 0.498206 |
| average | 1.000000 | 0.500000 | 0.500841 | 0.499895 |
| single | 1.000000 | 1.000000 | 1.000000 | 1.000000 |

Table 7. Precisions of the five clustering techniques when compared to the true and false labels.

| | weibo_stats | weibo_no_stats | twitter_stats | twitter_no_stats |
|---|---|---|---|---|
| k-means | 0.026163 | 0.264535 | 0.991611 | 0.243289 |
| ward | 0.000000 | 0.203488 | 0.996225 | 0.024329 |
| complete | 0.002907 | 0.598837 | 0.999581 | 0.932047 |
| average | 0.002907 | 0.997093 | 0.999581 | 0.999581 |
| single | 0.002907 | 0.002907 | 0.000419 | 0.000419 |

Table 8. Recalls of the five clustering techniques when compared to the true and false labels.

| | weibo_stats | weibo_no_stats | twitter_stats | twitter_no_stats |
|---|---|---|---|---|
| k-means | 0.050992 | 0.414579 | 0.663207 | 0.384871 |
| ward | NaN | 0.335731 | 0.665080 | 0.047502 |
| complete | 0.005797 | 0.508642 | 0.666760 | 0.649328 |
| average | 0.005797 | 0.666019 | 0.667320 | 0.666480 |
| single | 0.005797 | 0.005797 | 0.000839 | 0.000839 |

Table 9. F1 Scores of 5 Clustering Techniques Comparing to True and False Labels

Ultimately, it appears that using unsupervised clustering does not yield useful results when it comes to fake news truth detection in either Weibo or Twitter. The clusters, when using the true and false labels as an external validation index, produced better accuracy minutely than simple guess. This outcome could be anticipated when both Twitter and Weibo had no preference in number of clusters, suggesting the datasets do not cluster by a useful property.

## 5. Conclusion

In conclusion, through the comparison analysis, the observable characteristics of real and fake news does not show big difference between two different social media platforms despite China and USA being two differently governed nations. Looking at the feature importance analysis, the meaningful features when employing classification model had almost no difference between two social media as the most important feature is number of likes for both platforms. It indicates the practicability of implementing same classification model as a future reference. However, the optimal number of topics to be determined by LDA topic modelling varied a lot, which suggests more controlled and limited atmosphere in social media from authoritarian country.

Because of social media's innate characteristics of lack of censorship and freedom of speech, it is nearly impossible to intervene people's activity on social media platforms. One of the possible solutions that can be employed is developing a web application or labeling system within the social media platform to at least "inform" the users on accuracy and validity of the information they are consuming as a first step. It would help general users to get warnings on what facts they are engrossing and to be more cautious when absorbing thoughts from social media.

For the project, there are limitations that needs to be considered. First, limited accessibility to publicly available datasets with reliability in labeling is one of the restriction and problems to be solved as more studies are conducted. Especially, lack of reliable social media data is apparent for countries like China. Due to limited number of users compared to Twitter, the amount of data that can collected has clear perimeter of availability. Another limitation is specificity of datasets. Both datasets were chosen to cover COVID-19, however, this limited scope makes it difficult to find a clear difference in sentiments with regards to truth detection.

As a further development of current study, there are a few improvements that can be made. Application of more robust language models that make use of deep learning can be employed as it would allow texts to be interpreted as more than simple bag of words. Another improvement that would expand the applicability of the fake news comparison would be to include more languages, and to either use more accurate, possibly manual, translations or apply appropriate methods directly to the source languages.

## 6. References

Chu, S. K., Xie, R., & Wang, Y. (2021). Cross-language fake news detection. Data and Information Management, 5(1), 100–109. https://doi.org/10.2478/dim-2020-0025.

Das, S. D., Basak, A., & Dutta, S. (2021). A heuristic-driven ensemble framework for covid-19 fake news detection. arXiv preprint arXiv:2101.03545 .

Pang, H., Liu, J., & Lu, J. (2022). Tackling fake news in socially mediated public spheres: A comparison of Weibo and WeChat. Technology in Society, 70, 102025. https://doi.org/10.1016/j.techsoc.2022.102004

Reis, J. C., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019). Supervised learning for fake news detection. IEEE Intelligent Systems, 34, 76-81. doi: https://doi.org/ 10.1109/mis.2019.2899143

Shahi, G. K., Dirkson, A., & Majchrzak, T. A. (2021). An exploratory study of covid-19 misinformation on twitter. Online Social Networks and Media, 22, 100104.

Yang, C., Zhou, X., & Zafarani, R. (2021). Checked: Chinese covid-19 fake news dataset. Social Network Analysis and Mining (SNAM). doi: 10.1007/s13278-021-00766-8

## 7. Appendix

A.

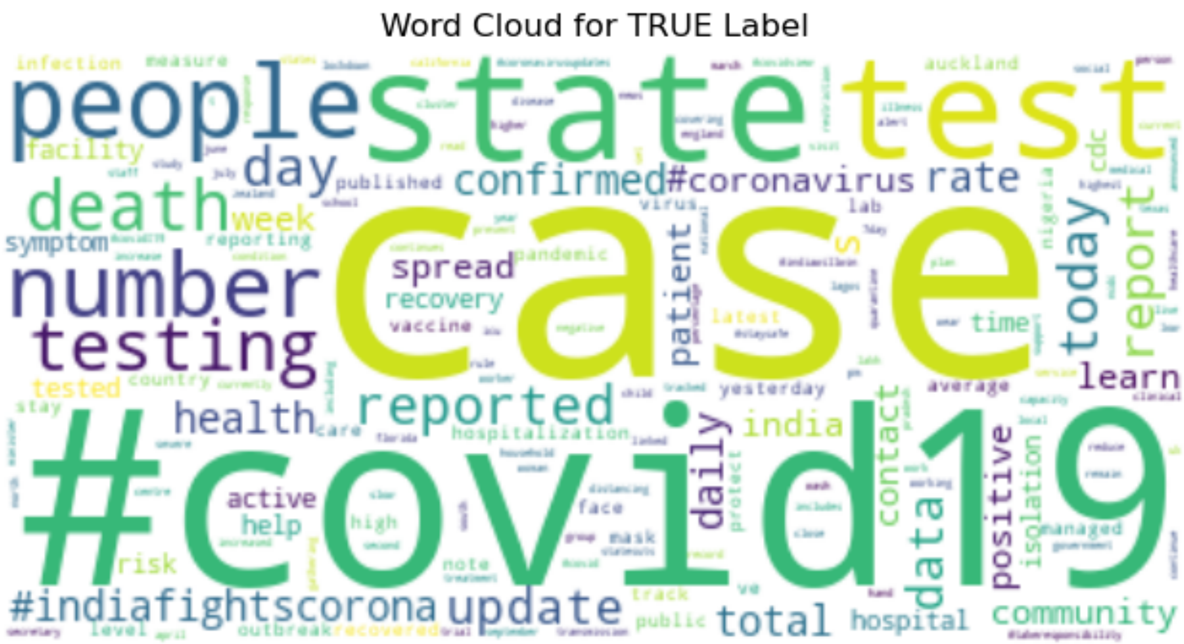



Figure 15. Word Clud of Tweets Labeled as True

B.


Figure 16. Word Cloud of Tweets Labeled as False

C.


Figure 17. Word Cloud of Weibo Labeled as "Fake"

D.


Figure 18. Word Cloud for Weibo Labeled as "Fake"

E.

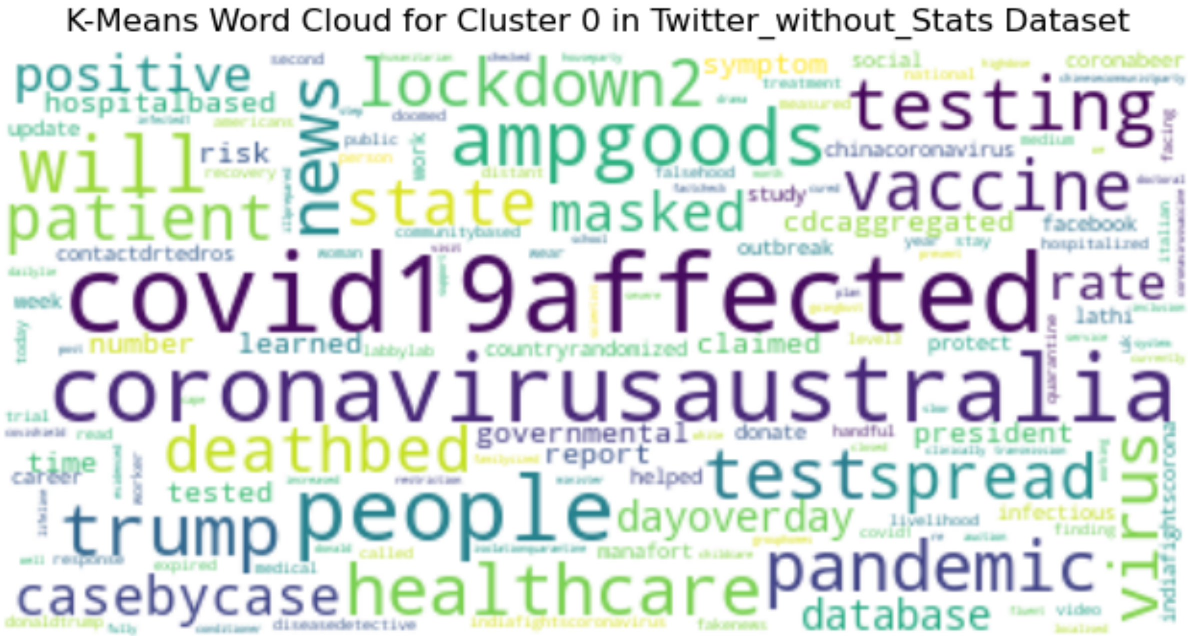



Figure 19. Word cloud for the first cluster of Twitter using k-means and no post statistics.

F.

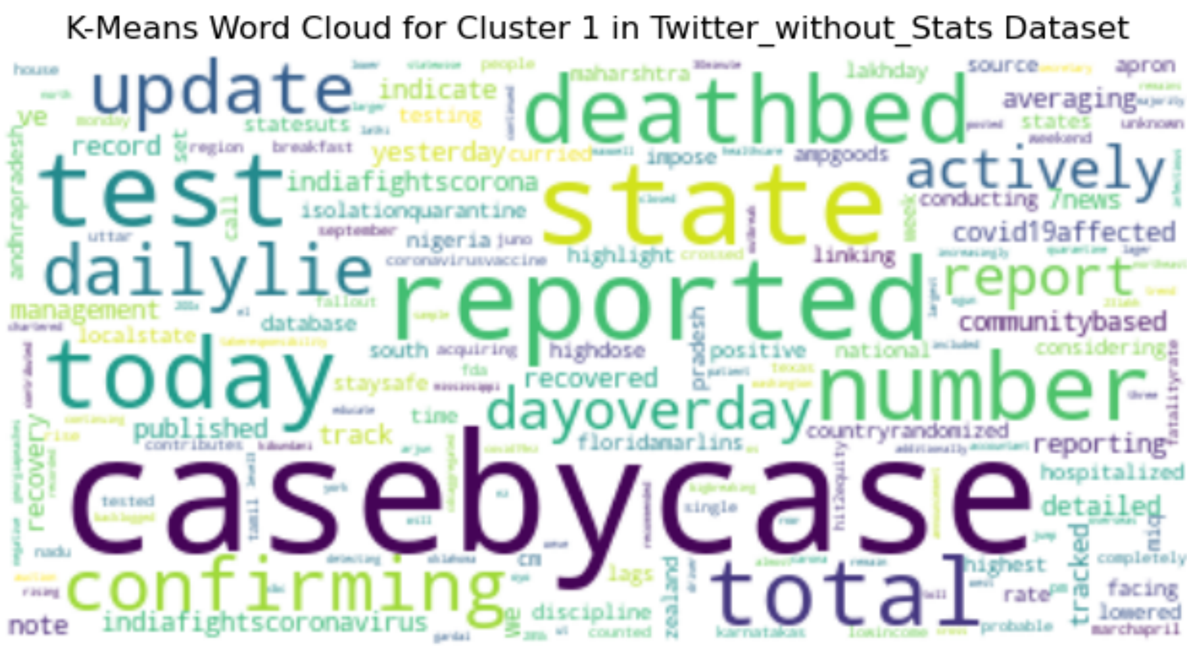



Figure 20. Word cloud for the second cluster of Twitter using k-means and no post statistics.

G.

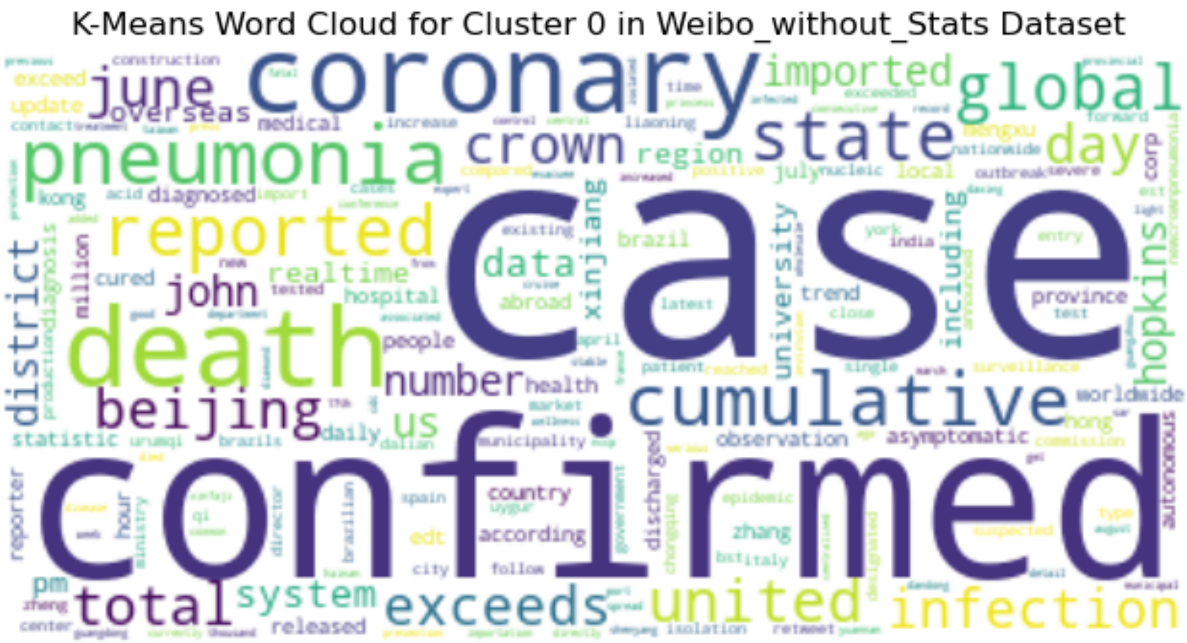



Figure 21. Word cloud for the first cluster of Weibo using k-means and no post statistics.

H.

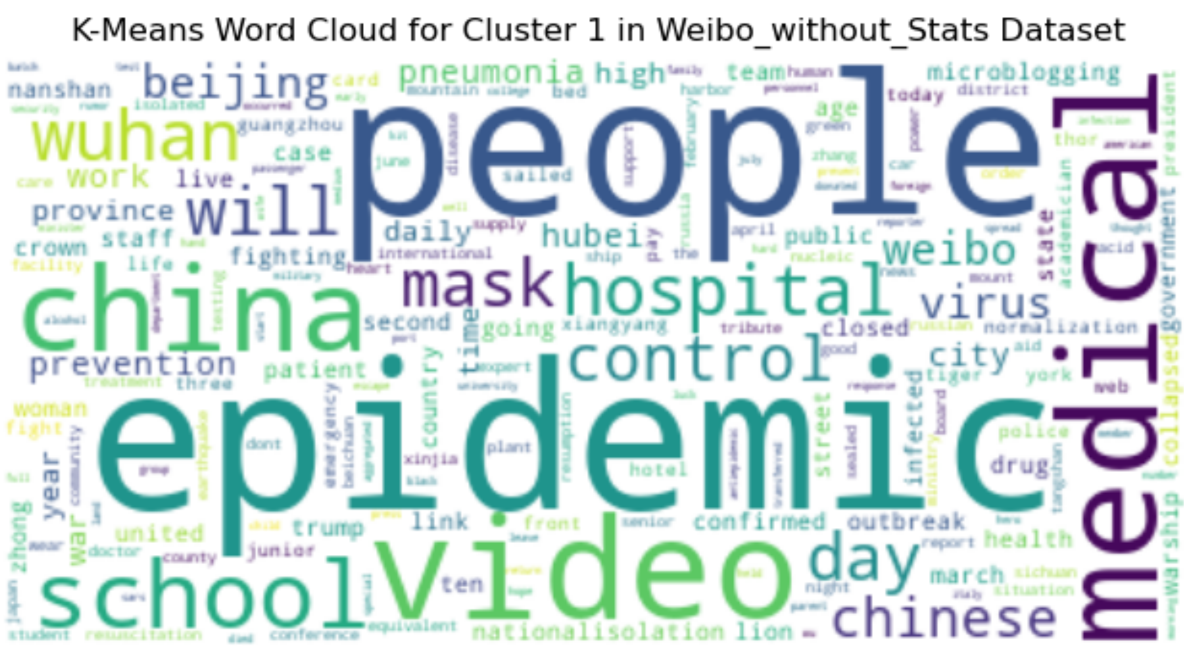



Figure 22. Word cloud for the second cluster of Weibo using k-means and no post statistics.