

Analysis of Fake News In Social Medias for Four Months during Lockdown in COVID-19

¹Shawni Dutta and ²Prof. Samir Kumar Bandyopadhyay

**¹Lecturer, Department of Computer Science, The Bhawanipur Education Society
College, Kolkata, India**

²Academic Advisor, The Bhawanipur Education Society College, Kolkata, India

Abstract

All over the world, development of micro blog and other social platform indicate that Social Media is now the focus and trend of the Internet. Daily life, study and work are influenced by news in Social Medias . Micro blog is new emergent type of media and it spreads information rapidly in the crowd in recent years. Suppose an user searches for specific information about one topic on micro blog. He/she found easily plenty of information related to his/her search in social medias. The problem is to find out the correct information.

Normally, multi-document summarization method deals with a collection of documents about one topic for extracting the valuable points and discards useless information. Actually, it needs to extract the topic content by adding topic factors and social patterns. Topic factor is the lexical information related to the topic. Social pattern relates to special interactive mode owned by online social network, such as comment and repost. People has been seen the fake news on mobile/internet during lockdown period. It is of no doubt that anyone with a social media account has seen at least one example of this.

Humanity's greatest challenges are to detect false information. Fake news are collected from 150 persons using social media The aim of the paper is to investigate the truthfulness of the news people share on social media using K-nearest Neighbour (KNN) based Classifier method.

Keywords: KNN, Fake News, , Social Media, Precision and COVID-19

Introduction

Fake news , if people believe, related to COVID-19 increases anxiety, stress, and even depression. The reason behind this is to stay-at-home, curfews, and closing of essential businesses. Many families are now phasing unemployment scenario. Some are forced to work from home due to the direction of companies. Many companies and businesses are working with loss due to coronavirus. This new life is stressful enough. So anxiety of people is clearly increased due to new life style [1].

Fake news is primarily spread via social media platforms, such as Facebook, Twitter, etc. Fake news is hard for some people to identify and can create confusion about true fact. People has doubt about accurate information. For example, When consumers learn information via fake news that the particular product is long-lasting. They have curiosity to see and purchase the product. Since it is not possible to purchase during lockdown or the information is incorrect. It made distress on consumers.

Researches have been going on to study the way fake news can be identified and detected[2]. It is true that the main way to combat fake news is by educating ourselves to understand fake news. Machine learning method may help to get identify false information during COVID-19. Facebook and Twitter are now taking measures to remove fake news and misinformation from their platforms [3]. The main objective of the study is to make sure that the information received from social media/mobile/internet are reliable and true. COVID-19 pandemic in 2020 dominates the media, both domestically and abroad. Alongside it is required for the attention on the pandemic first as well as to stop viral spread of fake news online related to coronavirus. At the time of COVID-19, the spread of fake news offers unique challenges and dangers to the public.

Machine learning approaches are applied for classification and recognizing fake news related to COVID-19. This paper requested 150 selected persons to address the problem of identifying fake news, people are requested to notify fake information in their social accounts. News in Mobile and email are also checked to get false news. Then machine learning based classifier is applied to analyse these fake news for measuring accuracy.

Related Works

Fake news is one of the serious issues in recent times addressed in the domain of COVID-19. People often post fake news in social media to mislead people regarding shopping of product with less price. It may guide other people to visit shopping mall to see the product. These news can be manipulated and hence it may mislead other people to visit the place. Natural Language Processing (NLP) has been used to detect fake news by extracting features[4-5]. Next, machine learning techniques are applied on these features. Lexicon based approaches may be one alternative to machine learning techniques that use dictionary or corpus to eliminate false news. Fake news in social media also hacked user accounts unknowingly. The fundamental study of fake news detection relies on three perspectives- how fake news is written, how fake news spreads and how a user is impressed by the fake news. Features related to news content and social context are extracted and a machine learning model is imposed to recognize fake news [6-7].

People share false claims about COVID-19 in part because they simply fail to think sufficiently about whether or not content is accurate. The news need to share with others without testing the news are correct or not. It can be of two types. Some participants are far worse at discerning between true and false content when they share on social media without knowing about accuracy of the fact. Participants engaged in more analytic thinking and had knowledge to separate fake news from the true news. They may sometimes discard the news and debarred the news spreading to social media. In this case, common sense is applied at the beginning of the study – i.e., asking people to judge the accuracy of a non-COVID-19-related news. It helps to stop spreading of false news in the social media. Some researchers found that these people can analyse false versus true news[4].

Data Set

The dataset contain news of 150 people obtaining from their social media accounts. These news are noted for analysis. The dataset are used in the proposed method from 20th March 2020 to 8th June 2020 for testing the overall performance of the approach. Before fitting this data to classifier, pre-processing is made on this dataset based on similarity news in social medias of 150 people. Pre-processing techniques include removal of not related to the corona, discard any communal news, deleted incomplete news. This prepares the dataset to be transformed into true features This features are the input to classifier.

Proposed Method and Results

The K-NN classifier is used for the value $k=5$ considering four months. Initially the classifier is trained actual data and is tested entire dataset for the prediction purpose. The performance measure metrics such as Accuracy, F1-score , and Cohen-Kappa score are used for evaluating the prediction. Accuracy identifies the ratio of true predictions over the total number of instances considered. However, the accuracy may not be enough for evaluating model's performance since it does not consider wrong predicted cases. If a fake news is treated as a true one, it creates a significant problem. Hence, it is necessary to consider false positive and false negative cases that compensate to misclassification. For measuring this compensation, precision and recall is quite necessary to be considered. Precision identifies the ratio of correct positive results over the number of positive results predicted by the classifier. Recall denotes the number of correct positive results divided by the number of all relevant samples. F1-Score is a parameter that is concerned for both recall and precision and it is calculated as the harmonic mean of precision and recall. Apart from all these measure, Cohen-Kappa Score is also considered for evaluation of the perfectness of the method. It is a statistical based measure that

finds out inter-rate agreement for qualitative items for classification problem. The results are shown from Table 1 to Table 5. From figure 1 to figure 5 shows the related graphs.

Days	Email	Mobile	WhatsAp	Twitter	Facebook
20	15	5	14	20	11
21	27	12	22	4	19
22	30	56	78	88	40
23	9	23	45	5	57
25	12	33	34	78	20
26	22	56	44	50	30
27	3	12	56	71	45
28	29	90	20	89	77
29	53	16	44	90	31
30	34	19	62	150	49
31	62	23	32	22	62

Table 1 March 2020 Data from Social Medias

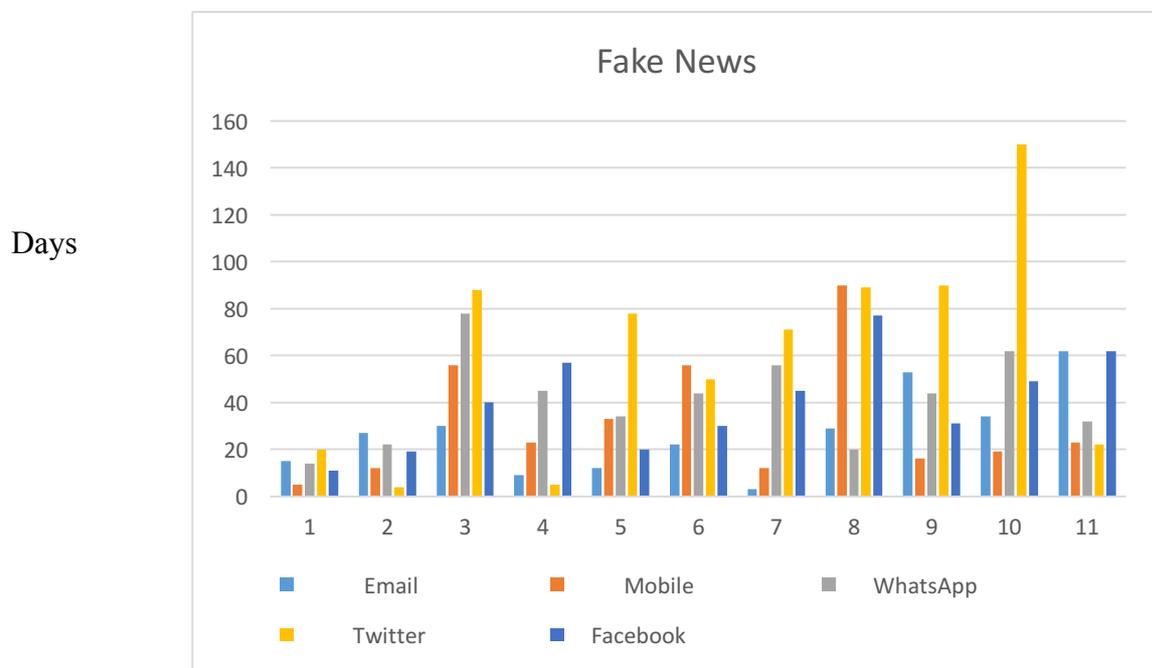


Figure 1 Fake News vs. Days for March 2020 of Social Medias Data

Days	Email	Mobile	Whatsapp	Twitter	Facebook
1	121	15	3	11	9
2	40	22	67	3	31
3	59	5	78	55	78
4	62	66	45	97	93
5	8	29	6	49	12
6	6	73	9	91	57
7	14	85	24	9	73
8	5	5	56	4	94
9	29	7	67	9	70
10	59	9	34	42	49
11	62	12	23	72	62
12	112	23	19	87	17
13	11	19	39	79	19
14	91	11	43	111	21
15	36	23	112	123	23
16	79	37	28	7	25
17	3	9	34	88	49
18	45	14	56	49	121
19	67	42	33	72	144
20	145	11	44	87	7
21	123	77	59	33	89
22	113	87	33	7	57
23	90	91	6	97	23
24	41	66	21	64	77
25	69	72	67	57	88
26	77	22	91	82	50
27	47	94	43	30	4
28	67	41	19	42	79
29	57	51	29	77	87
30	78	94	34	23	11

Table 2 April 2020 Data from Social Medias

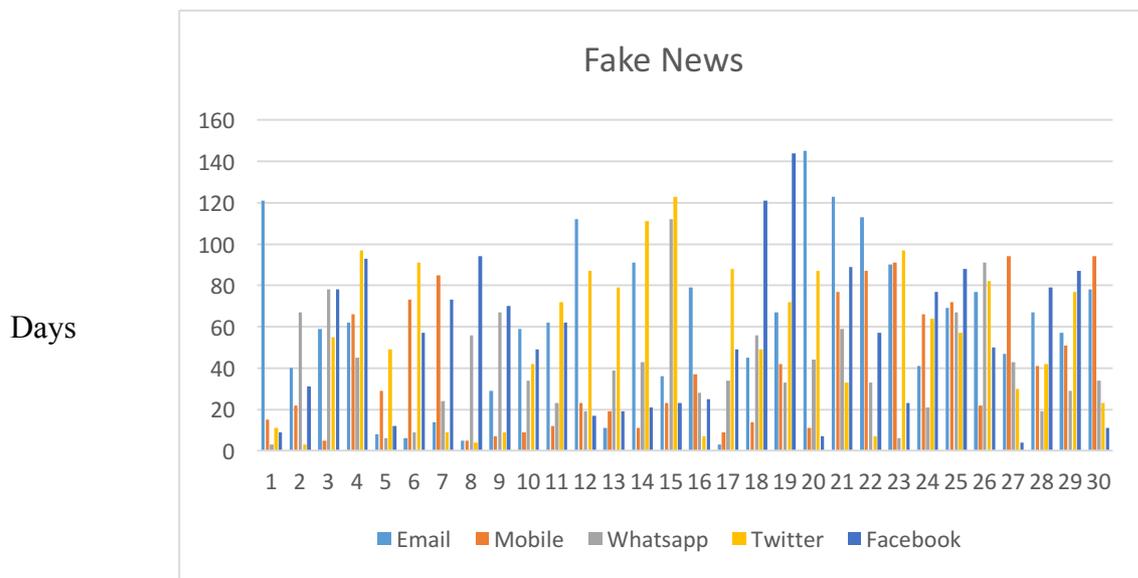


Figure 2 Fake News vs. Days for April 2020 of Social Medias Data

Days	Email	Mobile	Whatsapp	Twitter	Facebook
1	91	11	43	111	21
2	36	23	112	123	23
3	79	37	28	7	25
4	3	9	34	88	49
5	45	14	56	49	121
6	67	42	33	72	144
7	145	11	44	87	7
8	123	77	59	33	89
9	113	87	33	7	57
10	90	91	6	97	23
11	41	66	21	64	77
12	69	72	67	57	88
13	77	22	91	82	50
14	47	94	43	30	4
15	67	41	19	42	79
16	57	51	29	77	87
17	121	15	3	11	9
18	40	22	67	3	31
19	59	5	78	55	78
20	62	66	45	97	93
21	8	29	6	49	12
22	6	73	9	91	57
23	14	85	24	9	73
24	5	5	56	4	94
25	29	7	67	9	70
26	59	9	34	42	49
27	62	12	23	72	62
28	112	23	19	87	17
29	11	19	39	79	19
30	91	11	43	111	21
31	41	66	21	64	77

Table 3 May 2020 Data from Social Medias

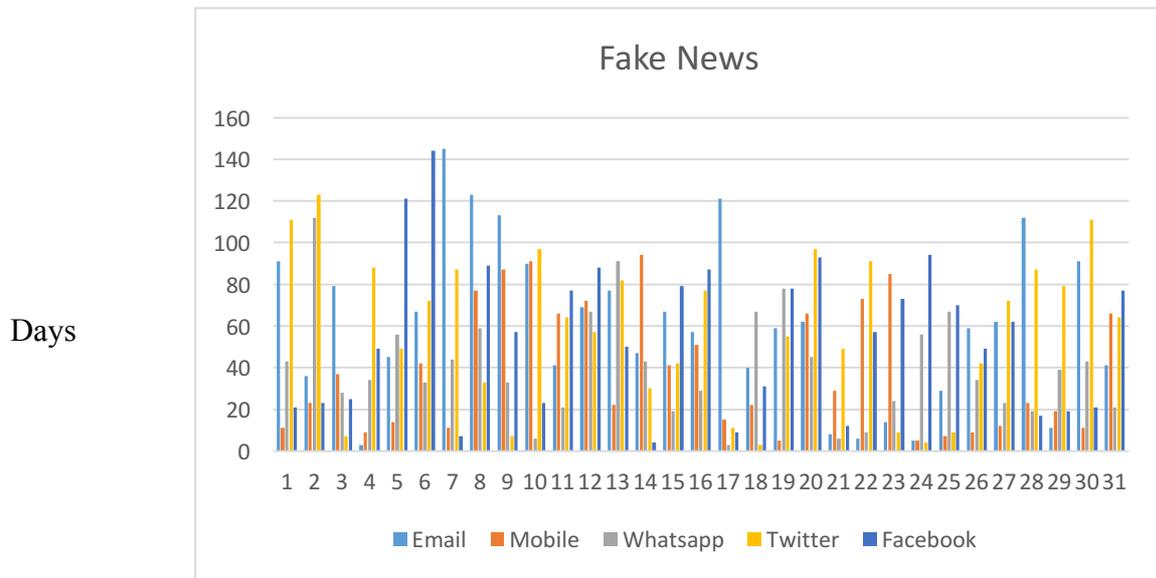


Figure 3 Fake News vs. Days for May2020 of Social Medias Data

June	Email	Mobile	Whatsapp	Twitter	Facebook
1	113	87	33	7	57
2	90	91	6	97	23
3	41	66	21	64	77
4	69	72	67	57	88
5	77	22	91	82	50
6	47	94	43	30	4
7	67	41	19	42	79
8	57	51	29	77	87

Table 4 June 2020 Data from Social Medias

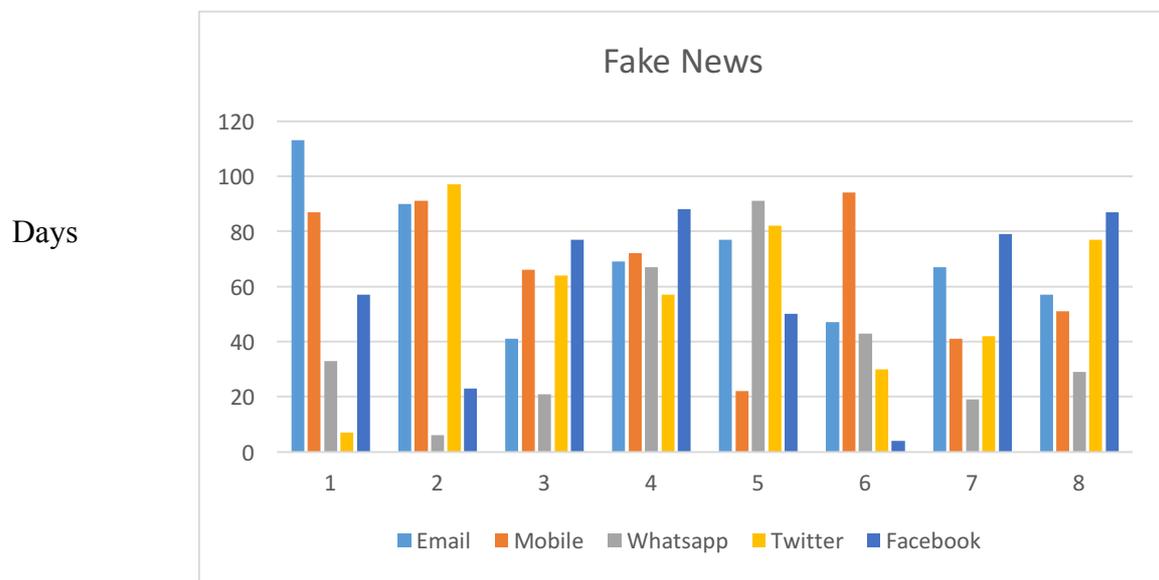


Figure 4 Fake News vs. Days for June 2020 of Social Medias Data

The Accuracy, F-score, and Cohen-Kappa score are used as performance measure for evaluating the prediction.

Month	Accuracy	F1-Score	Cohen-Kappa Score
March	76%	0.79	0,21
April	82%	0.88	0.29
May	89%	0.91	0.35
June	80%	0.93	0.33

Table 5 Performance Measure for Four Months during Lockdown

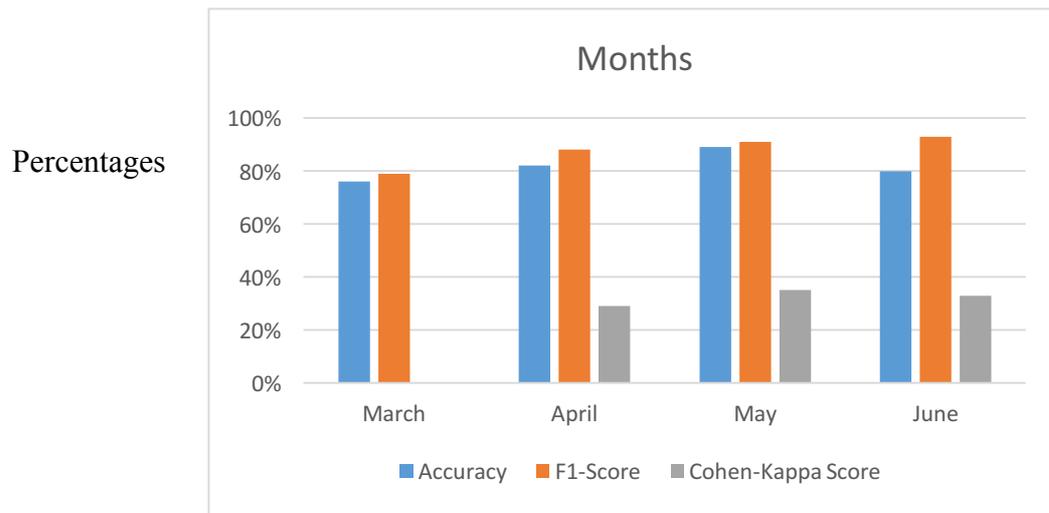


Figure 5 Performance Measure of Four Months

Conclusions

False news during COVID-19 is a great challenges remain to effective as long as it will continue. K-Nearest Neighbour Classifier is used as a study for measuring fake news in social medias during four months from March, 2020 to June 2020. The data set is created from 150 persons from their emails, mobiles, WhatsApp's and Twitters. The result obtained is quite interesting and shows the nature of false news during COVID-19.

References

1. Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. The covid-19 social media infodemic. arXiv preprint arXiv:2003.05004, 2020.
2. Lewandowsky, S. (2020). The "Post-Truth" World, Misinformation, And Information Literacy: A Perspective from Cognitive Science. In S. Goldstein (Ed.), *Informed Societies - Why Information Literacy Matters for Citizenship, Participation, and Democracy*. Facet.
3. McPhetres, J., & Pennycook, G. (2020). Science beliefs, political ideology, and cognitive sophistication. *PsyArXiv Working Paper*, 1–33. <https://doi.org/10.31219/OSF.IO/AD9V7>.
4. Miller, J. D. (1998). The measurement of civic scientific literacy. *Public Understanding of Science*, 7, 203–223. <https://doi.org/10.1088/0963-6625/7/3/001>
5. Mosleh, M., Pennycook, G., Arechar, A. A., & Rand, D. (2020). Digital Fingerprints of Cognitive Reflection. *PsyArXiv Working Paper*, 1–20.
6. Russonello, G. (2020, March 13). Afraid of Coronavirus? That Might Say Something About Your Politics. *The New York Times*.
7. Shivam Bansal (2020, February). [Real or Fake] Fake Job Posting Prediction, Version 1.