

Identification Of False Information Using Machine Learning

Ms. B. RAMA JYOTHI, Assistant Professor, Dept of CSE, KHIT, Guntur
B. DIVYA MADHURI, Kallam Haranadhareddy Institute of Technology, Guntur
G. JYOTHIRMAYEE, Kallam Haranadhareddy Institute of Technology, Guntur
D. BHARATH KUMAR, Kallam Haranadhareddy Institute of Technology, Guntur
M. KIRAN, Kallam Haranadhareddy Institute of Technology, Guntur
T. PUSHKAR, Kallam Haranadhareddy Institute of Technology, Guntur

ABSTRACT

In our modern era where the internet is ubiquitous, everyone relies on various online resources for news. Along with the increase in the use of social media platforms like Facebook, Twitter, etc. news spread rapidly among millions of users within a very short span of time. The spread of fake news has far-reaching consequences like the creation of biased opinions to swaying election outcomes for the benefit of certain candidates. Moreover, spammers use appealing news headlines to generate revenue using advertisements via click-baits. In this paper, we aim to perform binary classification of various news articles available online with the help of concepts pertaining to Artificial Intelligence, Natural Language Processing and Machine Learning. We aim to provide the user with the ability to classify the news as fake or real and also check the authenticity of the website publishing the news.

INTRODUCTION

These days' fake news is creating different issues from sarcastic articles to a fabricated news and plan government propaganda in some outlets. Fake news and lack of trust in the media are growing problems with huge ramifications in our society. Obviously, a purposely misleading story is "fake news" but lately blathering social media's discourse is changing its definition. Some of them now use the term to dismiss the facts counter to their preferred viewpoints. It is sought to produce a model that can accurately predict the likelihood that a given article is fake news. Facebook has been at the epic center of much critique following media attention. They have already implemented a feature to flag fake news on the site when a

user see's it ; they have also said publicly they are working on to to distinguish these articles *in* an automated way. Certainly, it is not an easy task. A given algorithm must be politically unbiased – since fake news exists on both ends of the spectrum – and also give equal balance to legitimate news sources on either end of the spectrum. In addition, the question of legitimacy is a difficult one. However, in order to solve this problem, it is necessary to have an understanding on what Fake News is. Later, it is needed to look into how the techniques in the fields of machine learning.

LITERATURE SURVEY

Mykhailo Granik et. al. in their paper shows a simple approach for fake news detection using naive Bayes classifier. This approach was implemented as a software system and tested against a data set of Facebook news posts. They were collected from three large Facebook pages each from the right and from the left, as well s three large mainstream political news pages (Politico, CNN, ABC News). They achieved classification accuracy of approximately 74%. Classification accuracy for fake news is slightly worse. This may be caused by the skewness of the dataset: only 4.9% of it is fake news.

Himank Gupta et. al. gave a framework based on different machine learning approach that deals with various problems including accuracy shortage, time lag (BotMaker) and high processing time to handle thousands of tweets in 1 sec. Firstly, they have collected 400,000 tweets from HSpam14 data set. Then they further characterize the 150,000 spam tweets and 250,000 non-spam tweets. They also derived some lightweight features along with the Top-30 words that are providing highest information gain from Bag-of-Words model. 4. They were able to achieve an accuracy of 91.65% and surpassed the existing solution by approximately 18%.

Marco L. Della Vedova et. al. first proposed a novel ML fake news detection method which, by combining news content and social context features, outperforms existing methods in the literature, increasing its accuracy up to 78.8%. Second, they implemented their method within a Facebook Messenger Chabot and validate it with a real-world application, obtaining a

fake news detection accuracy of 81.7%. Their goal was to classify a news item as reliable or fake; they first described the datasets they used for their test, then presented the content-based approach they implemented and the method they proposed to combine it with a social-based approach available in the literature. The resulting dataset is composed of 15,500 posts, coming from 32 pages (14 conspiracy pages, 18 scientific pages), with more than 2,300,000 likes by 900,000+ users. 8,923 (57.6%) posts are hoaxes and 6,577 (42.4%) are non-hoaxes.

PROPOSED

In Proposed work we are going to find the false information going viral on social and online media, or that being published by digital news media organization, may be monitored through the use of API's (Application Program Interface) of various social media platforms. This work assembling a data set of both fake and real news and employ a Linear model in order to create a model to classify news into fake or real based on its words and phrases.

RESULTS

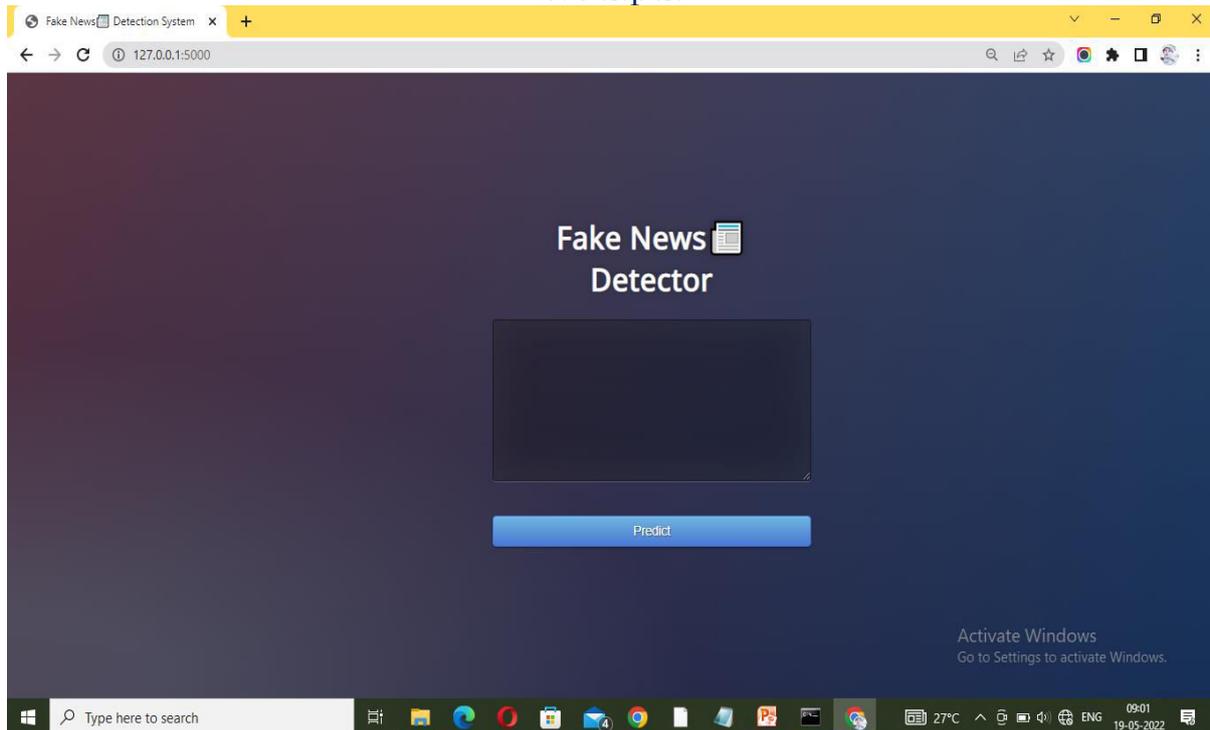
A	B	C	D
1	title	text	label
2	8476 You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fellow at the Freedom Center, is a	FAKE
3	10294 Watch The Exact Moment Paul Ryan Committed Political Suicide At A Trump Rally (VIDEO)	Google Pinterest Digg LinkedIn Reddit Stumbleupon Print Delicious Pocket	FAKE
4	3608 Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Monday that he will stop in Paris later	REAL
5	10142 Bernie supporters on Twitter erupt in anger against the DNC: 'We tried to warn you!'	â€” Kaydee King (@KaydeeKing) November 9, 2016 The lesson from tonight's	FAKE
6	875 The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners Hillary Clinton and Donald	REAL
7	6903 Tehran, USA		FAKE
8	7341 Girl Horrified At What She Watches Boyfriend Do After He Left FaceTime On	Share This Baylee Luciani (left), Screenshot of what Baylee caught on FaceTime	FAKE
9	95 â€œBritain's Schindlerâ€” Dies at 106	A Czech stockbroker who saved more than 650 Jewish children from Nazi Germa	REAL
10	4869 Fact check: Trump and Clinton at the 'commander-in-chief' forum	Hillary Clinton and Donald Trump made some inaccurate claims during an NBC	REAL
11	2909 Iran reportedly makes new push for uranium concessions in nuclear talks	Iranian negotiators reportedly have made a last-ditch push for more	REAL
12	1357 With all three Clintons in Iowa, a glimpse at the fire that has eluded Hillary Clinton's campaign	CEDAR RAPIDS, Iowa â€” had one of the most wonderful rallies of my	REAL
13	988 Donald Trump's Shockingly Weak Delegate Game Somehow Got Even Worse	Donald Trump's organizational problems have gone from bad to worse to	FAKE
14	7041 Strong Solar Storm, Tech Risks Today 50 News Oct.26.2016 [VIDEO]	Click Here To Learn More About Alexandra's Personalized Essences Psychic	FAKE
15	7623 10 Ways America Is Preparing For World War 3	October 31, 2016 at 4:52 am	FAKE
16	1571 Trump takes on Cruz, but lightly	Killing Obama administration rules, dismantling Obamacare and pushing through	REAL
17	4739 How women lead differently	As more women move into high offices, they often bring a style and approach	REAL
18	7737 Shocking! Michele Obama & Hillary Caught Glamorizing Date Rape Promoters	Shocking! Michele Obama & Hillary Caught Glamorizing Date Rape Promoters	FAKE
19	1876 Hillary Clinton in HUGE Trouble After America Noticed SICK Thing Hidden in this Picture... * LIBERTY WRITERS NEWS 0		FAKE
20	3304 What's in that Iran bill that Obama doesn't like?	Washington (CNN) For months, the White House and Congress have wrangled	REAL
21	3078 The 1 chart that explains everything you need to know about partisanship in America	While paging through Pew's best data visualizations of 2014 (it's awesome), I	REAL
22	2517 The slippery slope to Trump's proposed ban on Muslims	With little fanfare this fall, the New York developer who had planned to build	REAL
23	10348 Episode #160 â€œ Hail to the Deplorableâ€” with special guest Randy J	November 13, 2016 By 21wire Leave a Comment	FAKE
24	778 Hillary Clinton Makes A Bipartisan Appeal on Staten Island	Hillary Clinton told a Staten Island crowd today that she was the candidate who	REAL
25	3300 New Senate majority leader's main goal for GOP: Don't be scary	Mitch McConnell has an unusual admonition for the new Republican majority	REAL

a.Input

```
Select Command Prompt - python Fake_News_Det.py
19-05-2022 09:48 <DIR> .
19-05-2022 09:48 <DIR> ..
01-06-2021 10:37 1,307 Fake_News_Det.py
17-03-2022 10:18 18,723 Fake_News_Detection.ipynb
17-03-2022 10:17 489,898 model.pkl
17-03-2022 10:17 30,701,021 news.csv
01-06-2021 10:37 42 requirements.txt
17-03-2022 10:00 <DIR> static
01-06-2021 10:37 <DIR> templates
5 File(s) 31,210,991 bytes
4 Dir(s) 110,762,668,032 bytes free

C:\Users\Dell\Downloads\Fake_News_Detection>python Fake_News_Det.py
C:\Users\Dell\AppData\Local\Programs\Python\Python36\lib\site-packages\sklearn\base.py:315: UserWarning: Trying to unpickle estimator PassiveAggressiveClassifier from version 0.22.2.post1 when using version 0.24.2. This might lead to breaking code or invalid results. Use at your own risk.
UserWarning)
* Serving Flask app 'Fake_News_Det' (lazy loading)
* Environment: production
WARNING: This is a development server. Do not use it in a production deployment.
Use a production WSGI server instead.
* Debug mode: on
* Restarting with stat
C:\Users\Dell\AppData\Local\Programs\Python\Python36\lib\site-packages\sklearn\base.py:315: UserWarning: Trying to unpickle estimator PassiveAggressiveClassifier from version 0.22.2.post1 when using version 0.24.2. This might lead to breaking code or invalid results. Use at your own risk.
UserWarning)
* Debugger is active!
* Debugger PIN: 252-622-396
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

b. Output



CONCLUSIONS

In the 21st century, the majority of the tasks are done online. Newspapers that were earlier preferred as hard-copies are now being substituted by applications like Facebook, Twitter, and news articles to be read online. WhatsApp's forwards are also a major source. The growing problem of fake news only makes things more complicated and tries to change or hamper the opinion and attitude of people towards use of digital technology. When a person is deceived by the real news two possible things happen- People start believing that their perceptions about a particular topic are true as assumed. Thus, in order to curb the phenomenon, we have developed our Fake news Detection system that takes input from the user and classify it to be true or fake.

REFERENCES

- [1] C. Anagnostopoulos, T. Alexandropoulos, V. Loumos and E. Kayafas, "Intelligent traffic management through MPEG-7 vehicle flow surveillance," in *Proceedings of IEEE International Symposium on Modern Computing*, 2006, pp. 202-207.
- [2] N. Mani, and B. Srinivasan, "Application of artificial neural network model for optical character recognition," presented at the *IEEE International Conference on Systems, Man, and Cybernetics*, 1997.
- [3] R. P. van Heerden, E. C. Botha, *Optimization of Vehicle License Plate Segmentation and Symbol Recognition*, Department of Electrical, Electronic and Computer engineering, University of Pretoria, South Africa, 2010.
- [4] F. M. Rodriguez, X. F. Hermida, *New Advances in Automatic Reading of V.L.P.s (Vehicle License Plates)*, *Proceedings in SPC- 2000 (Signal Processing and Communications)*, Marbella, September 2000.
- [5] F.M. Rodriguez, M. G. Saburido, J. L. A. Castro, *New Methods for Automatic Reading of V.L.P.s (Vehicle License Plates)*, *SPPRA- 2002 (Signal Processing Pattern Recognition and Applications)*, June 2002.
- [6] A. E. Savakis, *Adaptive Document Image Thresholding Using foreground and Background Clustering*, *International Conference on Image Processing*, October 1998.
- [7] Ondrej Martinsky, *Algorithmic and mathematical Principles of automatic number plate Recognition systems*, Brno University of technology, 2007.
- [8] Sushruth Shastry, Gunasheela G, Thejus Dutt, Vinay D S and Sudhir Rao Rupanagudi, i - *A novel algorithm for Optical Character Recognition (OCR)*, *IEEE* 2013.
- [9] W. Badawy, "Automatic License Plate Recognition (ALPR): A State of the Art Review", 2012
- [10] Sukhpreet Singh *Optical Character Recognition Techniques: A Survey*, *Journal of Emerging Trends in Computing and Information Sciences*, Vol. 4, No. 6 June 2013.
- [11] T. Naito, T. Tsukada, K. Yamada, K. Kozuka, and S. Yamamoto, *Robust licenseplate recognition method for passing vehicles under outside environment*, *Trans. Veh. Technol.*, vol. 49, no. 6, pp. 23092319, Nov. 2000.
- [12] M. Usman Akram, Zabeel Bashir, Anam Tariq and Shoab A Khan, *Geometric Feature Points Based Optical Character Recognition*, *IEEE Symp. Industrial Elec. & App.*, Sept. 2013.
- [13] Dr Savita Gael and Savita Dabas, *Vehicle Registration Plate Recognition System Using Template Matching*, *IEEE 97B-1 -4799-1 607-B/13*, 2013.