Fake and Real News detection Using Python

Fakhra Akhtar[1], Faizan Ahmed Khan[2], Muhammad Tauseef Hanif[2]

[1]Faculty of Computer Science an Information Technology, RIPHA University Lahore, Pakistan

[2]COMSATS University Islamabad, Lahore-Campus, Pakistan

ABSTRACT

In the digital age, fake news has become a well-known phenomenon. The spread of false evidence is often used to confuse mainstream media and political opponents, and can lead to social media wars, hatred arguments and debates.Fake news is blurring the distinction between real and false information, and is often spread on social media resulting in negative views and opinions. Earlier Research describe the fact that false propaganda is used to create false stories on mainstream media in order to cause a revolt and tension among the masses The digital rights foundation DRF report, which builds on the experiences of 152 journalists and activists in Pakistan, presents that more than 88 % of the participants find social media platforms as the worst source for information, with Facebook being the absolute worst. The dataset used in this paper relates to Real and fake news detection. The objective of this paper is to determine the Accuracy , precision , of the entire dataset .The results are visualized in the form of graphs and the analysis was done using python.

The results showed the fact that the dataset holds 95% of the accuracy. The number of actual predicted cases were 296. Results of this paper reveals that The accuracy of the model dataset is 95.26 % the precision results 95.79 % whereas recall and F-Measure shows 94.56% and 95.17% accuracy respectively.Whereas in predicted models there are 296 positive attributes , 308 negative attributes 17 false positives and 13 false negatives. This research recommends that authenticity of news should be analysed first instead of drafting an opinion, sharing fake news or false information is considered unethical journalists and news consumers both should act responsibly while sharing any news.

Keywords- Fake News Detection, Python, Accuracy

I. INTRODUCTION

Fake news is a piece of information that consists of planned misinformation, and it is circulated through some medium, especially social media. The unverified information is spread through traditional news platforms, in the pursuit of some agendas, concerted efforts are designed to draft something out of nothing. Whenever we think about fake news and misleading information we consider mainstream media and social network. After making necessary research on the defined topic we consider and built up any opinion, regarding fake & real. This propaganda is being circulating around for many decades, today internet is the latest source of information with the single click you can see a massive information regarding any topic. Today in the era of tech revolution Things are moving in a different direction, fighting an utterly different war, and on an entirely different level. This fake news is controlled by various masses and business tycoons to fulfill political agenda and they are often perceived as one of the greatest threats to democracy, independence and national interests.

Facebook, the most popular social media platform, claims to have more than 1.23 billion daily active users as of December 2016 (Facebook 2017). While it started as a site through which we can share personal ideas and updates with friends, a survey carried out in the United States found that 44 percent of the population get their news from Facebook (Gottfried and Shearer 2016). Social media sites are not only marked by having a mass audience, they also facilitate speedy exchange and spread of information. Unfortunately, they have also facilitated the spread of wrong information.

The first phase of the paper explains about literature review regarding to the defined dataset i.e. fake news detection The second phase deals with the methodology data analysis tool. The third part discusses the research results based on the detailed analysisFourth phase deals with Discussion, conclusion and references.

II. LITERATURE REVIEW

(1)Balmau at el (2018). In their research describe the fact that Today's social media platforms enable to spread both authentic and fake news very quickly. They developed a mechanism to limit the spread of fake news which is not based on content, by using a Bayesian approach they estimated the trustworthiness of future news items they also evaluated the effectiveness and overhead of this technique on a large Twitter graph. They identified that more than 99% of fake news items with no false positives. The performance impact is very small: the induced overhead on the 90th percentile latency is less than 3%, and less than 8% on the throughput of user operations.(2) Brody at el (2018) In their research consider fake news as a danger to democracy. According to them until now there has been no clear understanding of how to define fake news, and how to model it Their paper addresses both these issues, .they used two approaches for the modelling of fake news and its impact in elections an referendums. The

first approach is based on the idea of a representative voter, is shown to be suitable to obtain a qualitative understanding of phenomena associated with fake news at a macroscopic level. Whereas The second approach, based on the idea of an election microstructure, describes the collective behaviour of the electorate by modelling the preferences of individual voters. Results of their study shows that the mere knowledge that pieces of fake news may be in circulation goes a long way towards mitigating the impact of fake news.(3) Monti at.el (2019). In their research consider social media as one of the main news sources for millions of people around the globe due to their low cost, and easy They used core algorithms of classical convolutional neural networks to graphs, activity, social graph, and news propagation. Their model was trained and tested on news stories, verified by professional fact-checking organizations, that were spread on Twitter. Their results showed that social network structure and propagation are important features allowing highly accurate Secondly they observed that fake news can be reliably detected a an early stage, after just a few hours of propagation. Third, they tested the aging of their model on training and testing data separated in time. (4) Vicario, at.el (2019) in their research introduced a framework for promptly identifying polarizing content on social media and, thus, predicting future fake news topics. They validated the performances of the proposed methodology on amassive Italian Facebook dataset, showing that they are able to identify topics that are susceptible to misinformation with 77% accuracy. They also able to recognize fake news with 91% accuracy. Their results concluded the fact that a series of characteristics related to users' behavior on online social media such as Facebook is an important step towards the mitigation of misinformation phenomena by supporting the identification of potential misinformation targets and thus the design of tailored counter-narratives. (5) Kleinberg at. el (2017) in their

research describe the fact that the production of misleading information in everyday access media outlets such as social media feeds, news blogs, and online newspapers have made it challenging to identify trustworthy news sources, thus increasing the need for computational tools able to provide insights into the reliability of online content. They are basically focusing on the automatic identification of fake content in online news. They introduced two novel datasets for the task of fake news detection, covering seven different news domains. They also described the collection, annotation, and validation process in detail and presented several exploratory analyses on the identification of linguistic differences in fake and legitimate news content. Secondly, they conducted a set of learning experiments to build accurate fake news detectors. In addition, they provided a comparative analyses of the automatic and manual identification of fake news. Furthermore, they also addressed the task of automatic identification of fake news. They introduced two new fake news datasets, one obtained through crowdsourcing and covering six news domains, and another one obtained from the web covering celebrities. They developed classification models that rely on a combination of lexical, syntactic, and semantic information, as well features representing text readability properties. The results of their model showed that they Our best performing achieved accuracies that are comparable to human ability to spot fake content. (6) Wang at. el (2017) In their research consider Social media for news consumption as a double-edged sword. On the one hand, its low cost, easy access, and rapid dissemination of information lead people to seek out and consume news from social media. On the other hand, it enables the wide spread of fake news", i.e., low quality news with intentionally false information. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society. Therefore, fake news detection on social media has recently become an emerging research that is attracting tremendous attention. They also describe the fact that Fake news detection on social media presents unique characteristics and challenges that make existing detection algorithms from traditional news media incentive or not applicable. Fake news are intentionally written to mislead readers to believe false information, which makes it difficult and nontrivial to detect based on news content; They included auxiliary information, such as user social engagements on social media, to help make a determination. Second, exploiting this auxiliary information is challenging in and of itself as users' social engagements with fake news produce data that is big, incomplete, unstructured, and noisy. Because the issue of fake news detection on social media is both challenging and relevant, we conducted this survey to further facilitate research on the problem. They presented a comprehensive review of detecting fake news on social media, including fake news characterizations on psychology and social theories, existing algorithms from a data mining perspective, evaluation metrics and representative datasets. They also discussed related research areas, open problems, and future research directions for fake news detection on social media.

III. METHODOLOGY

Artificial intelligence is considered as a use of computerized reasoning (AI) It centers around the advancement of computer programs that can get to information and use it for themselves for the purpose of decision making in this research the dataset is interpreted and analyzed using Python.

The Data

The dataset used in this research paper is denoted by news.csv. It has a shape of 7796×4. The first column identifies the news, the second and third are the title

and text, and the fourth column has labels denoting whether the news is REAL or FAKE

PYTHON

Python is a suitable language for both learning and real world programming. It is a powerful high-level, object-oriented programming language created by Guido van Rossum and is also known to have a great quantity of libraries that assist with data analysis & scientific computing. It is a very high-level language and has many sources for learning. It supports a wide variety of third party tools which makes it much easier to use and motivates the users to continue with, Having an elegant syntax. It's much easier to read and write Python programs compared to other languages like: C++, Java, C#. it makes programming fun and allows you to focus on the solution rather than syntax. Furthermore, Python is also more forgiving of errors, so you'll still be able to compile and run your program until you hit the problematic part. Python is a flexible, simple coding programming language. This language can support different styles of programming including structural and object-oriented.

IV. RESULTS

In order to interpret the result for the defined data set python programming tool was used. the results are shown in the form of various tables and bar chart

I.   Table 1 Dataset evaluation result

Results	Values
Accuracy	95.26814
Precision	95.79288
Recall	94.56869
F-Measure	95.17685

Table 1 shows the evaluation results applied on the defined dataset. The accuracy of the model dataset is

95.26 % the precision results 95.79 % whereas recall and F-Measure shows 94.56% and 95.17% accuracy respectively.

**Table 2** predicted class results

Model		Prediction Yes	Prediction No
Actual	Yes	296	17
Class	No	13	308

Table 2 shows the predicted class results so with this model we have 296 positive attributes, 308 negative attributes 17 false positives and 13 false negatives.

Table 3 actual and fake news

Model	Predicted class		
Actual	Yes	TP	FN
Class	No	FP	TN

Table 3 represents actual and fake news in the form of prediction noted by True Positive (TP)
& false positive (FP)
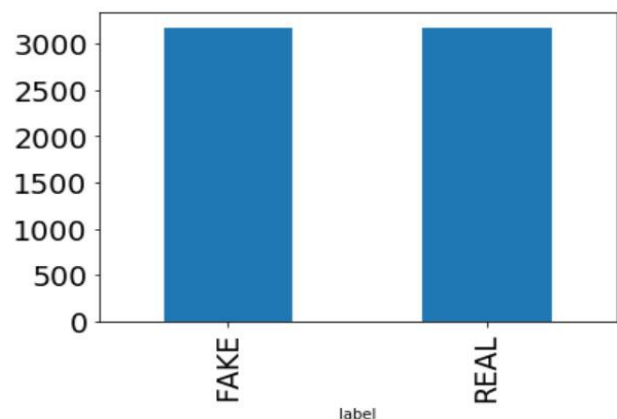
Figure 1. Bar Chart

Graphical representation of Real & Fake News

DISCUSSION

Based on results it is observed that The accuracy of the model dataset is 95.26 % the precision results 95.79 % whereas recall and F-Measure shows 94.56% and 95.17% accuracy respectively. Whereas in predicted models there are 296 positive attributes, 308 negative attributes 17 false positives and 13 false negatives. In bar chart the results of the real and fake news are shown.

V. CONCLUSION AND RECOMMENDATION

The results show that The accuracy of the model dataset is 95.26 % the precision results 95.79 % whereas recall and F-Measure shows 94.56% and 95.17% accuracy respectively. Whereas in predicted models there are 296 positive attributes, 308 negative attributes 17 false positives and 13 false negatives. This research recommends that authenticity of news should be analyzed first instead of drafting an opinion, sharing fake news or false information is considered unethical journalists and news consumers both should act responsibly while sharing any news. The public and the media are equally responsible for this problem. People take advantage of the power they hold, and small drops of misinformation turn into a huge sea of misunderstandings.

VI. ACKNOWLEDGMENT

VII. REFERENCES

[1]. Balmau, O., Guerraoui, R., Kermarrec, A. M., Maurer, A., Pavlovic, M., & Zwaenepoel, W. (2018). Limiting the Spread of Fake News on Social Media Platforms by Evaluating Users' Trustworthiness. arXiv preprint arXiv:1808.09922

[2]. Brody, D. C., & Meier, D. M. (2018). How to model fake news. arXiv preprint arXiv:1809.00964.

[3]. Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M. M. (2019). Fake news detection on social media using geometric deep learning. arXiv preprint arXiv:1902.06673.

[4]. Vicario, M. D., Quattrociocchi, W., Scala, A., & Zollo, F. (2019). Polarization and fake news: Early warning of potential misinformation targets. ACM Transactions on the Web (TWEB), 13(2), 1-22.

[5]. Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news" A typology of scholarly definitions. Digital journalism, 6(2), 137-153.

[6]. Volkova, S., Shaffer, K., Jang, J. Y., & Hodas, N. (2017, July). Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 647-653).

[7]. Apampa, O. (2019). Fake News: A Data-Driven Analysis Of The Role Of Influencers In The Spread Of Misinformation And Disinformation On Selected Social Media Platforms.

[8]. Rügenhagen, M., Beck, T. S., & Sartorius, E. J. (2020). Information Integrity in the Era of Fake News.

[9]. TIOBE Software Index (2011). "TIOBE Programming Community Index Python". 1

[10]. "Programming Language Trends - O'Reilly Radar". Radar.oreilly.com. 2 August 2006.

[11]. Bakir, V., & McStay, A. (2018). Fake news and the economy of emotions: Problems, causes, solutions. Digital journalism, 6(2), 154-175.

[12]. Gelfert, A. (2018). Fake news: A definition. Informal Logic, 38(1), 84-117.

[13]. Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of fake news by social bots. arXiv preprint arXiv:1707.07592, 96, 104.

[14]. Conroy, N. K., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1), 1-4.

[15]. Jang, S. M., & Kim, J. K. (2018). Third person effects of fake news: Fake news regulation and media literacy interventions. Computers in Human Behavior, 80, 295-302.

[16]. Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., ... & Gao, J. (2018, July). Eann: Event adversarial neural networks for multi-modal fake news detection. In Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining (pp. 849-857).

[17]. Waisbord, S. (2018). Truth is what happens to news: On journalism, fake news, and post-truth. Journalism studies, 19(13), 1866-1878.

[18]. Nielsen, R. K., & Graves, L. (2017). " News you don't believe": Audience perspectives on fake news.

[19]. Rubin, V. L., Conroy, N., Chen, Y., & Cornwell, S. (2016, June). Fake news or truth? using satirical cues to detect potentially misleading news. In Proceedings of the second workshop on computational approaches to deception detection (pp. 7-17).

[20]. Rochlin, N. (2017). Fake news: belief in post-truth. Library hi tech.

[21]. Kanchan Mahajan, Proff.J.S.Chitode, "Waste Bin Monitoring

Cite this article as :