

Improved Accuracy for Fake News in Social Media Using Logistic Regression Comparing Naive Bayes Classifier

K.Jeevan^a and K.V.Kanimozhi^{b,1}

^aResearch Scholar, Department of CSE, Saveetha School of Engineering,

^bProject Guide, Department of CSE, Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamilnadu, India

Abstract. The fundamental purpose of the work is to detect the Fake News in Social Media with the use of Machine Learning Algorithms. TRUE and FAKE dataset is used to detect false news. This dataset carries the record of data i.e. TRUE or FAKE news. Fake News detection is accomplished via Logistic Regression and Naive Bayes classifier. Naive-Bayes algorithm is a simple approach mainly used for classification. Sample size has been determined to be 20 for both the groups using G Power 80%. Logistic Regression algorithm provides mean accuracy of 97.5% when compared to Naive Bayes algorithm with mean accuracy of 89.43%. Statistical significance value is obtained as 0.002 ($p < 0.05$). Logistic Regression has extensively higher accuracy than Naive Bayes algorithm.

Keywords. Machine Learning, Artificial Neural Network, Novel Logistic Regression, Naive Bayes, Fake news detection.

1. Introduction

In recent times, Fake news grabbed the utmost attention of many researches across the globe. Fake news is any story or a context that is used to manipulate the public to believe in the news that is fake[1]. It is mentioned that around 2 million people are misled due to the fake in the year of 2019 [2]. A most recent report by a news blog that declared online platforms accounted for 50% of news that are fake[3]. Almost 62% of the world is reading the news from online platforms which makes them believe that Fake news which can be harmful will have a major impact on human life. Fake news detection provides a detailed insight of a news based on which the genuineness of the news is detected[4]. Many researchers have carried out research on Fake information Detection with the use of ML techniques and 20 associated works in IEEE Digital explore and 10 articles were posted in research gate. used naive bayes and decision tree algorithms to predict the Fake News Severity[5]. [6] predict the incidence of false information with the use of logistic regression and SVM with 91% accuracy to predict that accuracy achieved for logistic regression is higher than the support vector

¹ K.V.Kanimozhi, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, E.Mail.kanimozhikv.sse@saveetha.com

machines. An analysis on fake news proposed that Artificial Neural Network and logistic regression and achieved an accuracy of 88% and discussed that 72% of people in the world are using social media platforms for reading the news which include facebook, instagram etc[7]. Analyzing the key elements mainly detecting the fake information using classification and regression models using Artificial neural networks[8].

Development of the recall and precision of hybrid combinations of different methods, Decision tree, XGBoost and the results depict that Logistic regression showed higher accuracy than the other algorithms[9]. By using naive bayes algorithm accuracy around 84% has achieved. Our wide portfolio in research has translated into publications in numerous interdisciplinary projects.[10-14]. Now we are focusing on this topic. Based on the above literature survey, traditional Classification Algorithms have lesser classification accuracy along with poor stability. Inaccuracy of false information may lead to many difficult consequences for the people. Therefore this paper focuses on more accurate results in detecting the fake news with improved accuracy.

2. Materials and Methods

The work is carried out in the Computer Vision Laboratory, Saveetha School of Engineering. The quantity of groups recognized to analyze is two. Group one is Logistic Regression and Group two is Naive Bayes[15]. 'TRUE' and 'FAKE' datasets are downloaded from the Kaggle Data World. This dataset contains the information of data i.e. TRUE or FAKE news. Total of 1500 rows of data is used for training and validating the data. News from different regions all over the world at different dates are taken for this research work. Dividing input dataset to train and test out components in which 80% of statistics is utilized to train and 20% is utilized for test algorithms.

Using Pandas Library, the dataset is read into the program and the data in this dataset is given as input. The given data is processed using both methods and the records are splitted into train and test groups. Using these Logistic Regression and Naive Bayes Algorithms that are imported from sklearn and sequential libraries and from pipeline () function, Fake news is detected. The output is displayed either as True or Fake. Naive Bayes is one of the simplest and most efficient algorithms in machine learning which gives accurate results and quick predictions. This algorithm also has some vectorization techniques like count vectorizer, TFIDF transformer etc. which are used in converting the data that is given into the encoded format and then the data will be trained. The software program device that is used to analyze the Logistic Regression and Naive Bayes algorithm is Python Programming Language in Jupyter notebook.

2.1 Statistical Analysis:

Analysis tool used is IBM SPSS. Independent sample t test is carried out and significance values are obtained. The independent variables are Text, Subject and the dependent variable is Target details the effect as True or Fake.

Table 1. Accuracy of Fake News Detection classification using Logistic Regression (mean accuracy=97.5%, mean loss=2.5%)

Test	Accuracy	Loss
Test 1	98.23	1.77
Test 2	99.86	0.14
Test 3	93.23	6.77
Test 4	96.5	3.5
Test 5	100.0	0.0
Test 6	100.0	0.0
Test 7	99.9	0.1
Test 8	92.58	7.42
Test 9	100.0	0.0
Test 10	97.37	2.63

3. Result

In this study, false information detection is done using Logistic Regression and it is proved that Novel Logistic Regression has shown better significant results and accuracy than Naive Bayes Algorithm. Accuracy is measured for more than a few samples with the use of both Algorithms as given in Table 1 and 2. To gain the correlation between the accuracy values, SPSS tool was used which gave descriptive statistical analysis with values for Mean, Standard error mean and Standard Deviation for the two algorithms Logistic Regression and Naive Bayes Classifier is shown in Table 3. Mean Accuracy value for Logistic Regression and Naive Bayes Algorithm is 97.5267 and 89.4324 respectively. Mean Loss value for Logistic Regression and Decision Tree is 2.4843 and 10.6545 respectively. Table 4 exhibits the independent T test analysis of both Algorithms. Depends on the evaluation there is a considerable distinction between two groups having the p value of 0.002 ($p < 0.05$).

Table 2. Accuracy of fake news detection classification using Naive Bayes algorithm

(Mean Accuracy=89.43%, mean Loss= 10.57%)

Test	Accuracy	Loss
Test 1	83.33	16.67
Test 2	85.71	14.29
Test 3	75.0	25.0
Test 4	100.0	0.0
Test 5	100.0	0.0
Test 6	91.66	8.34
Test 7	87.5	12.5
Test 8	96.0	4.0
Test 9	77.77	22.23
Test 10	81.23	18.77

Table 3. Group Statistics analysis for both algorithms based on Accuracy and Loss.

	Groups	N	Mean	Std. Deviation	Std.Error Mean
Accuracy	Logistic Regression	10	97.5267	2.23016	.40837
	Naive Bayes	10	89.4324	5.65063	.74476
	Logistic Regression	10	2.4843	2.58886	.40933
Loss	Naive Bayes	10	10.6545	4.71026	.74476

Table 4. An Independent sample T-Test detecting Fake information with the use of Logistic Regression and Naive Bayes algorithms. Logistic regression algorithm seems to function significantly higher than Naive Bayes algorithm (p=0.002)

		Levene' s test for Equality of Variances		T-test for equality of means						
		F	Sig.	t	df	Sig.(2-tailed)	Mean Difference	Std. Error Difference	95% Confidence interval of the difference	
									Lower	Upper
Accuracy	Equal variances assumed	9.82	.002	9.38	78	<.001	7.97	.8493	6.18212	9.66146
	Equal variances not assumed			9.38	60.5	<.001	7.97	.8493	6.16592	9.66920
Loss	Equal variances assumed	9.74	.002	-9.37	78	<.001	- 7.97	.8498	-9.6573	-6.2736
	Equal variances not assumed			-9.37	60.59	<.001	-7.96	.8498	-9.6650	-.26592

Logistic Regression and Naive Bayes Algorithms provide the accuracy of 97.5% and 89.43% respectively. Loss value for Logistic Regression and Decision Tree algorithm is 2.5 and 11.55 respectively. Standard deviation of the Novel LR algorithm (2.23) is barely higher than SVM (5.65).

4. Discussion

In this study, accuracy of the Novel Logistic Regression algorithm is significantly higher than the Naive Bayes Algorithm. From all the previous literature surveys, it is shown that the Novel Logistic Regression algorithm produced better accurate results. By using Logistic Regression algorithm by processing false news containing the data of false information and predicted the fake information with an accuracy rate of 82%[8]. By carrying out False information detection and determined that Logistic regression is the most desirable algorithm for detect false information when a minimal wide variety of attributes are taken into consideration[2] .

5. Conclusion

An efficient and accurate model is developed for detecting the fake news using a Logistic regression algorithm. The proposed model shows significantly higher accuracy than Naive bayes. By using Logistic Regression and Naive Bayes algorithm and the accuracy obtained using Logistic regression is 77%. The issue of proposed method is fake news detection is in general categorized the use of solely three foremost attributes. In future it is necessary to consist of greater attributes which have an affect on the detection of fake news

References

- [1] Shu, Kai, and Huan Liu. Detecting Fake News on Social Media. Morgan & Claypool Publishers. 2019.
- [2] Tiwari, Vanya, Ruth G. Lennon, and Thomas Dowling. Not Everything You Read Is True! Fake News Detection Using Machine Learning Algorithms.” 2020 31st Irish Signals and Systems Conference (ISSC). <https://doi.org/10.1109/issc49989.2020.9180206>. 2020.
- [3] Kudarvalli, Harika, and Jinan Fiaidhi. Detecting Fake News Using Machine Learning Algorithms.” <https://doi.org/10.36227/techrxiv.12089133.v1>. 2020.
- [4] Waikhom, Lilapati, and Rajat Subhra Goswami. Fake News Detection Using Machine Learning. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.3462938>. 2019.
- [5] Han, Wenlin, and Varshil Mehta. Fake News Detection in Social Networks Using Machine Learning and Deep Learning: Performance Evaluation. 2019 IEEE International Conference on Industrial Internet (ICII). <https://doi.org/10.1109/icii.2019.00070>. 2019.
- [6] Sharma, Vishal. 2020. “Machine Learning Algorithms via Detection of Fake News. International Journal for Research in Applied Science and Engineering Technology. <https://doi.org/10.22214/ijraset.2020.6125>.
- [7] Masood, Razan, and Ahmet Aker. The Fake News Challenge: Stance Detection Using Traditional Machine Learning Approaches.” Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. <https://doi.org/10.5220/0006898801280135>. 2018
- [8] Baarir, Nihel Fatima, and Abdelhamid Djeflal. Fake News Detection Using Machine Learning. 2nd International Workshop on Human-Centric Smart Environments for Health and Well-Being (IHSH). <https://doi.org/10.1109/ihsh51661.2021.9378748>. 2021.
- [9] Kaliyar, Rohit Kumar, Anurag Goswami, and Pratik Narang. Multiclass Fake News Detection Using Ensemble Machine Learning. 2019 IEEE 9th International Conference on Advanced Computing (IACC). <https://doi.org/10.1109/iacc48062.2019.8971579>. 2019
- [10] Shu, Kai, Suhang Wang, Dongwon Lee, and Huan Liu. Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities. Springer Nature. 2020
- [11] Jain, Ravindra Kumar, Sridhar Prem Kumar, and W. S. Manjula. Comparison of Intrusion Effects on Maxillary Incisors among Mini Implant Anchorage, J-Hook Headgear and Utility Arch. Journal of Clinical and Diagnostic R.2014; 8 (7) eSearch: JCDR: ZC21–24.

- [12] Venu, Harish, Lingesan Subramani, and V. Dhana Raju. Emission Reduction in a DI Diesel Engine Using Exhaust Gas Recirculation (EGR) of Palm Biodiesel Blended with TiO₂ Nano Additives. *Renewable Energy* . 2019; September(140): 245–63.
- [13] Nandhini, J. S. Thaslima, K. Yuvaraj Babu, and Karthik Ganesh Mohanraj. Size, Shape, Prominence and Localization of Gerdy's Tubercle in Dry Human Tibial Bones. *Journal of Advanced Pharmaceutical Technology & Research*. 2018; 11 (8): 3604.
- [14] Kannan, Roghith, and M. S. Thenmozhi. Morphometric Study of Styloid Process and Its Clinical Importance on Eagle's Syndrome." *Journal of Advanced Pharmaceutical Technology & Research*. 2016; 9 (8): 1137.
- [15] Zhou, Y., and A. Kumar. Human Identification Using Palm-Vein Images. *IEEE Transactions on Information Forensics and Security*. 2011; 6 (4): 1259–74.
- [16]
- [17] Baarir, Nihel Fatima, and Abdelhamid Djeflal. Fake News Detection Using Machine Learning. 2nd International Workshop on Human-Centric Smart Environments for Health and Well-Being (IHSH). <https://doi.org/10.1109/ihsh51661.2021.9378748>.2021
- [18] Babar, Akshada, Nalini Jagtap, Akshata Mithari, Aakash Shukla, and Prachi Chaudhari. A Survey on Fake News Detection Techniques and Using a Blockchain Based System to Combat Fake News. *International Journal of Computer Applications*. <https://doi.org/10.5120/ijca2020920291>.2020.
- [19] Granik, Mykhailo, and Volodymyr Mesyura. Fake News Detection Using Naive Bayes Classifier. 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON). <https://doi.org/10.1109/ukrcon.2017.8100379>.2017.
- [20] Han, Wenlin, and Varshil Mehta. Fake News Detection in Social Networks Using Machine Learning and Deep Learning: Performance Evaluation. 2019 IEEE International Conference on Industrial Internet (ICII). <https://doi.org/10.1109/ici.2019.00070>.2019.
- [21] Jain, Ravindra Kumar, Sridhar Prem Kumar, and W. S. Manjula. Comparison of Intrusion Effects on Maxillary Incisors among Mini Implant Anchorage, J-Hook Headgear and Utility Arch. *Journal of Clinical and Diagnostic Research: JCDR*.2014; 8 (7): ZC21–24.
- [22] Kaliyar, Rohit Kumar, Anurag Goswami, and Pratik Narang. Multiclass Fake News Detection Using Ensemble Machine Learning.2019 IEEE 9th International Conference on Advanced Computing (IACC). <https://doi.org/10.1109/iacc48062.2019.8971579>.2019.
- [23] Kannan, Roghith, and M. S. Thenmozhi. Morphometric Study of Styloid Process and Its Clinical Importance on Eagle's Syndrome." *Journal of Advanced Pharmaceutical Technology & Research*. 2016; 9 (8): 1137.
- [24] Konagala, Varalakshmi, and Shahana Bano. Fake News Detection Using Deep Learning. *Research Anthology on Fake News, Political Warfare, and Combatting the Spread of Misinformation*. <https://doi.org/10.4018/978-1-7998-7291-7.ch025>.2021.
- [25] Kudarvalli, Harika, and Jinan Faiidhi. Detecting Fake News Using Machine Learning Algorithms." <https://doi.org/10.36227/techrxiv.12089133.v1>. 2020.
- [26] Masood, Razan, and Ahmet Aker. The Fake News Challenge: Stance Detection Using Traditional Machine Learning Approaches." *Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*. <https://doi.org/10.5220/0006898801280135>. 2018.
- [27] Nandhini, J. S. Thaslima, K. Yuvaraj Babu, and Karthik Ganesh Mohanraj. Size, Shape, Prominence and Localization of Gerdy's Tubercle in Dry Human Tibial Bones. *Journal of Advanced Pharmaceutical Technology & Research*. 2018; 11 (8): 3604.
- [28] Sharma, Vishal. 2020. "Machine Learning Algorithms via Detection of Fake News. *International Journal for Research in Applied Science and Engineering Technology*. <https://doi.org/10.22214/ijraset.2020.6125>.
- [29] Shu, Kai, and Huan Liu. *Detecting Fake News on Social Media*. Morgan & Claypool Publishers. 2019.
- [30] Shu, Kai, Suhang Wang, Dongwon Lee, and Huan Liu. *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities*. Springer Nature. 2020.
- [31] Tiwari, Vanya, Ruth G. Lennon, and Thomas Dowling. Not Everything You Read Is True! Fake News Detection Using Machine Learning Algorithms." 2020 31st Irish Signals and Systems Conference (ISSC). <https://doi.org/10.1109/issc49989.2020.9180206>. 2020.
- [32] Venu, Harish, Lingesan Subramani, and V. Dhana Raju. Emission Reduction in a DI Diesel Engine Using Exhaust Gas Recirculation (EGR) of Palm Biodiesel Blended with TiO₂ Nano Additives. *Renewable Energy* . 2019; September(140): 245–63.
- [33] Waikhom, Lilapati, and Rajat Subhra Goswami. Fake News Detection Using Machine Learning. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3462938>. 2019.
- [34] Zhou, Y., and A. Kumar. Human Identification Using Palm-Vein Images. *IEEE Transactions on Information Forensics and Security*. 2011; 6 (4): 1259–74.