



FAKE NEWS DETECTION IN CHATTING APPLICATION WITH RANDOM FOREST OVER LSTM

Kasturi pogiri¹, S.John Justin Thangaraj^{2*}

Article History: Received: 12.12.2022

Revised: 29.01.2023

Accepted: 15.03.2023

Abstract:

Aim: The aim of the proposed research is to develop false news detection using the Random forest model and improve accuracy with neural networks in contrast to the long short term memory model. **Materials and Methods:** The random forest model is applied on data, which is a text file containing sequences, a collection of words LSTM for predicting the accuracy of fake news that compares two sources. Model of long-term short-term memory. It has been suggested and developed to have random forests. The size of the sample The G Power value of 0.8 was used to calculate the number of people in each category. The precision was excellent. Random forest (56 percent) was the most effective in spotting bogus news. When compared to LSTM, the least mean error is (40%). **Results:** The accuracy was maximum in detecting the fake news in social media using random forest 56% with long short term memory model 40% for the same dataset. **Conclusion:** The study proves that random forest exhibits better accuracy than long short term memory in detecting the fake news on e-news applications.

Keywords: Machine Learning, Innovative detection, Random forest, LSTM, Naive Bayes Classifiers, Deep Neural Networks.

¹Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

^{2*}Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

1. Introduction

The use of misinformation in American politics drew a lot of attention in 2016, especially after Donald Trump was elected. The phrase "fake news" has become widely used in this industry, particularly to describe the misleading and inaccurate pieces published in order to profit from web page traffic using (Kavanagh and Rich 2018). Most academics nowadays are concentrating their efforts on developing a model capable of accurately predicting whether a given article is true or false news using advanced machine learning techniques. It is vital to determine and objectively define what makes the new site "legitimate". The main aim is to develop false news detection using the Random forest model and improve accuracy with neural networks in contrast to the long short term memory model

Over the past 5 years, on spam detection using machine learning, 18,400 articles have been published in Google Scholar, 27 journal papers are available in IEEE Xplore, 1,773 articles are available in ScienceDirect. The adverse effects of false info will lead people to assume that Hillary Clinton has a foreign child, with the aim of attracting readers that President Trump is planning to repeal the first amendment in order to kill India's crowds due to false stories spread through WhatsApp (Geary, n.d.). People believe what they read on websites or social media today and do not verify to determine if the information supplied is accurate or fake (Klee 2015). Manually differentiating between fake and real news is challenging since people must spend a significant amount of time analyzing news references and ensuring their accuracy (Mazarr et al. 2019) to avoid the difficulty machine learning algorithms and artificial intelligence like naive bias algorithm is utilized. As a result, there is a significant demand for an automatic and intelligent model for detecting fake news. As a result, the detection of false news attracts a lot of attention from researchers all over the world (Kavanagh and Rich 2018). In Singapore, Google and Facebook opposed the introduction of new regulations to combat fake news, stating that existing legislation is adequate to address the problem and that training people on how to distinguish between fake and true news is an effective approach to combating fake news using Naive Bayes Classifiers. Despite all of the efforts made by the existing society, people, technology, and processes, fake news continues to exist in some form or another on a daily basis. To eradicate fake news some innovative detention methods are introduced from artificial intelligence and machine learning algorithms like, Accuracy, Naive Bayes Classifiers, Deep Neural Networks. Our team has extensive knowledge and research experience that

has translated into high quality publications (Pandiyar et al. 2022; Yaashikaa, Devi, and Kumar 2022; Venu et al. 2022; Kumar et al. 2022; Nagaraju et al. 2022; Karpagam et al. 2022; Baraneedharan et al. 2022; Whangchai et al. 2022; Nagarajan et al. 2022; Deena et al. 2022)

Naïve Bayes Classifier is an innovative detection algorithm as it helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. The Naive Bayes Classifier is a part of a deep neural network. A deep neural network (DNN) is an artificial neural network (ANN) with multiple layers between the input and output layers. There are different types of neural networks but they always consist of the same components: neurons, synapses, weights, biases, and functions. The main aim of the proposed research is to develop false news detection using the Random forest model and improve accuracy with neural networks in contrast to the long short term memory model

2. Materials and Methods

This study was implemented at Data analytics lab, Saveetha School of Engineering, SIMATS using jupyter notebook software, and hardware configurations are intel i3 core processor, 64GB HDD, 4GB RAM, and the software configurations are windows OS, python jupyter notebook. The work was carried out on 6328 records from the text file from online dataset kaggle website. The accuracy in predicting the next word was performed by evaluating two groups. A total 150 epochs were performed on each group to achieve better accuracy. The study uses a dataset downloaded from kaggle website.

Random Forest

Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. Random forest is straightforward. Import the python libraries required for the fake news detection. The pseudocode for Random forest are

Step 1: import libraries

Step2: import dataset

Step3: creating the target column

Step4: concatenating the title text of the news

Step5: converting data column to data time format

Step6: appending two data sets

LSTM

LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown

duration between important events in a time series. Import the python libraries required for the fake news detection. The pseudocode for LSTM are:

Step 1: import libraries

Step2: import dataset

Step3: creating the target column

Step4: concatenating the title text of the news

Step5: converting data column to data time format

Step6: appending two data sets

Statistical analysis

The SPSS statistical software was used in the research for statistical analysis. Variables like test data are independent whereas predicted data is dependent on test data. Group statistics and independent sample tests were performed on the experimental results and the graph was built for two graphs with two parameters under the study. The analysis of the experiment is represented in bar graph (comparison between random forest and LSTM algorithm). A table for comparison of loss of accuracy is drawn from the spss tool. The above analysis paves a path to conclude the effectiveness of the algorithm and final conclusion is drawn.

3. Results

The proposed random forest technique and the existing LSTM algorithm were run in a jupyter notebook one at a time. The accuracy and loss values of random forest and LSTM increase as the sample sets are run for a number of iterations. Table 1 shows the significant levels for random forest and $P=0.01$ was used to evaluate LSTM models. With a 40.50 percent chance of being correct, both random forest and LSTM have a less significant level less than 0.05.

4. Discussion

Observations were conducted among the study groups randomforest and LSTM by varying sample size, from observations the proposed LSTM performed better in terms of detecting fake news by achieving the accuracy and less error rate compared to the LSTM (Lewis, Choudhury, and Chitty 2015). The challenges identified as determinant causes of incorrect information, as well as various correlations between news items, writers, and sources subjects (Rocha et al. 2021). This paper provides a novel automatic model for detecting false positives (Choy et al. 2005). Information trustworthiness It creates a highly dispersed environment. a framework for analyzing news stories interpretations, On the basis of a collection of authors and subjects at the same time Textual material generates explicit and implicit feature vectors (Li et al. 2021). The term "false information" refers to the type of daily mail being sent out. False data or fake stories are displayed on purpose. Both traditional print media and new

media are being used to spread the word. Online social networks. For the objective of false information detection, the approach uses Deep Learning algorithms (Li et al. 2021; Fort, Nicolàs-Aragó, and Palacín 2021). The rapid growth of erroneous media production and distribution creates an immediate necessity for certain altered news reports to be tagged and discovered in real time. Accurate false information analysis is difficult to obtain since it requires a way for recognising complications in natural language. False news is a term that is used by both traditional and non-traditional news sources, such as social media, to describe false news or advertising that offers misleading information (Yip et al. 2018). The basic purpose for propagating this information is to confuse viewers, harm some individuals' reliability, or benefit from headlines (Preston et al. 2021).

Although the proposed methodology attained satisfactory results, the limitation in the proposed approach is that there needs to be improved accurate news detection. In future this can be combined with more data text files which can produce better results.

5. Conclusion

The results show that the proposed random forest outperforms LSTM in terms of accuracy and loss for fake news detection. The proposed random forest proves with better accuracy (54%) when compared with LSTM for detecting fake news in chatting applications.

Declarations

Conflict of Interests

No conflict of interests in this manuscript.

Authors contributions

Author PK was involved in data collection, data analysis implementation, algorithm forming and manuscript writing. Author JJT was involved in designing the workflow, guidance and review of manuscript.

Acknowledgements

We would like to acknowledge saveetha school of engineering, saveetha institute of medical and technical sciences (formerly known as saveetha university) for providing facilities and constant services to this study.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Vee Eee Technologies and Solutions Pvt.Ltd, Chennai
2. Saveetha university

3. Saveetha Institute of Medical and Technical Sciences

4. Saveetha School of Engineering

6. REFERENCES

- Baraneedharan, P., Sethumathavan Vadivel, C. A. Anil, S. Beer Mohamed, and Saravanan Rajendran. 2022. "Advances in Preparation, Mechanism and Applications of Various Carbon Materials in Environmental Applications: A Review." *Chemosphere*. <https://doi.org/10.1016/j.chemosphere.2022.134596>.
- Choy, E. H. S., C. Smith, C. J. Doré, and D. L. Scott. 2005. "A Meta-Analysis of the Efficacy and Toxicity of Combining Disease-Modifying Anti-Rheumatic Drugs in Rheumatoid Arthritis Based on Patient Withdrawal." *Rheumatology* 44 (11): 1414–21.
- Deena, Santhana Raj, A. S. Vickram, S. Manikandan, R. Subbaiya, N. Karmegam, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2022. "Enhanced Biogas Production from Food Waste and Activated Sludge Using Advanced Techniques – A Review." *Bioresource Technology*. <https://doi.org/10.1016/j.biortech.2022.127234>.
- Fort, Joana, Adrià Nicolàs-Aragó, and Manuel Palacín. 2021. "The Ectodomains of rBAT and 4F2hc Are Fake or Orphan α -Glucosidases." *Molecules* 26 (20). <https://doi.org/10.3390/molecules26206231>.
- Geary, Lacie. n.d. "Spread of False News Stories on Facebook: An Assessment of Credibility Cues and Personality." <https://doi.org/10.33915/etd.5654>.
- Karpagam, M., R. Beulah Jeyavathana, Sathiya Kumar Chinnappan, K. V. Kanimozhi, and M. Sambath. 2022. "A Novel Face Recognition Model for Fighting against Human Trafficking in Surveillance Videos and Rescuing Victims." *Soft Computing*. <https://doi.org/10.1007/s00500-022-06931-1>.
- Kavanagh, and Michael D. Rich. 2018. *Truth Decay: An Initial Exploration of the Diminishing Role of Facts and Analysis in American Public Life*. Rand Corporation.
- Klee, Miles. 2015. "True False." <https://doi.org/10.2307/j.ctt207g8bt>.
- Kumar, P. Ganesh, P. Ganesh Kumar, Rajendran Prabakaran, D. Sakthivadivel, P. Somasundaram, V. S. Vigneswaran, and Sung Chul Kim. 2022. "Ultrasonication Time Optimization for Multi-Walled Carbon Nanotube Based Thermol-55 Nanofluid: An Experimental Investigation." *Journal of Thermal Analysis and Calorimetry*. <https://doi.org/10.1007/s10973-022-11298-4>.
- Lewis, Celine, Mahrufa Choudhury, and Lyn S. Chitty. 2015. "'Hope for Safe Prenatal Gene Tests'. A Content Analysis of How the UK Press Media Are Reporting Advances in Non-Invasive Prenatal Testing." *Prenatal Diagnosis* 35 (5): 420–27.
- Li, Lan, Aisha Aldosery, Fedor Vitiugin, Naomi Nathan, David Novillo-Ortiz, Carlos Castillo, and Patty Kostkova. 2021. "The Response of Governments and Public Health Agencies to COVID-19 Pandemics on Social Media: A Multi-Country Analysis of Twitter Discourse." *Frontiers in Public Health* 9 (September): 716333.
- Mazarr, Michael J., Ryan Michael Bauer, Abigail Casey, and Luke J. Matthews. 2019. *The Emerging Risk of Virtual Societal Warfare: Social Manipulation in a Changing Information Environment*.
- Nagarajan, Karthik, Arul Rajagopalan, S. Angalaeswari, L. Natrayan, and Wubishet Degife Mammo. 2022. "Combined Economic Emission Dispatch of Microgrid with the Incorporation of Renewable Energy Sources Using Improved Mayfly Optimization Algorithm." *Computational Intelligence and Neuroscience* 2022 (April): 6461690.
- Nagaraju, V., B. R. Tapas Bapu, P. Bhuvaneswari, R. Anita, P. G. Kuppusamy, and S. Usha. 2022. "Role of Silicon Carbide Nanoparticle on Electromagnetic Interference Shielding Behavior of Carbon Fibre Epoxy Nanocomposites in 3-18GHz Frequency Bands." *Silicon*. <https://doi.org/10.1007/s12633-022-01825-1>.
- Pandiyan, P., R. Sitharthan, S. Saravanan, Natarajan Prabakaran, M. Ramji Tiwari, T. Chinnadurai, T. Yuvaraj, and K. R. Devabalaji. 2022. "A Comprehensive Review of the Prospects for Rural Electrification Using Stand-Alone and Hybrid Energy Technologies." *Sustainable Energy Technologies and Assessments*. <https://doi.org/10.1016/j.seta.2022.102155>.
- Preston, Stephanie, Anthony Anderson, David J. Robertson, Mark P. Shephard, and Narisong Huhe. 2021. "Correction: Detecting Fake News on Facebook: The Role of Emotional Intelligence." *PloS One* 16 (10): e0258719.
- Rocha, Yasmim Mendes, Gabriel Acácio de Moura, Gabriel Alves Desidério, Carlos Henrique de Oliveira, Francisco Dantas Lourenço, and Larissa Deadame de Figueiredo Nicolette. 2021. "The Impact of Fake News on Social Media and Its Influence on Health during the COVID-19 Pandemic: A Systematic Review." *Zeitschrift Fur Gesundheitswissenschaften = Journal of*

Public Health, October, 1–10.
 Venu, Harish, Ibham Veza, Lokesh Selvam, Prabhu Appavu, V. Dhana Raju, Lingesan Subramani, and Jayashri N. Nair. 2022. “Analysis of Particle Size Diameter (PSD), Mass Fraction Burnt (MFB) and Particulate Number (PN) Emissions in a Diesel Engine Powered by Diesel/biodiesel/n-Amyl Alcohol Blends.” *Energy*.
<https://doi.org/10.1016/j.energy.2022.123806>.
 Whangchai, Niwooti, Daovieng Yaibouathong, Pattranan Junluthin, Deepanraj Balakrishnan, Yuwalee Unpaprom, Rameshprabu Ramaraj, and Tipsukhon Pimpimol. 2022. “Effect of

Biogas Sludge Meal Supplement in Feed on Growth Performance Molting Period and Production Cost of Giant Freshwater Prawn Culture.” *Chemosphere* 301 (August): 134638.
 Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. “Advances in the Application of Immobilized Enzyme for the Remediation of Hazardous Pollutant: A Review.” *Chemosphere* 299 (July): 134390.
 Yip, S. Y., D. Namah, R. Cook, and C. Isles. 2018. “It Must Be True ... I Read It in the Tabloids.” *The Journal of the Royal College of Physicians of Edinburgh* 48 (3): 251–56.

Tables and Figures

Table 1: Comparison of accuracy and loss obtained between Randomforest and LSTM

	Algorithm	N	mean	Std.deviation	std.Error mean
Accuracy	RandomForest	3	40.2433	16.66648	9.62239
	LSTM	3	30.333	10.0664	5.81187
Loss	Randomforest	3	33.8667	12.00181	6.92925
	LSTM	3	51.8733	3.39837	1.96205

Table 2:

Accuracy	Independent Samples Test								
	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
Equal variances assumed	0.267	0.611	4.901	18	0.438	0.7790	0.15894	0.44509	1.11291
Equal variances not assumed			4.901	17.201	0.439	0.7790	0.15894	0.44397	1.11403

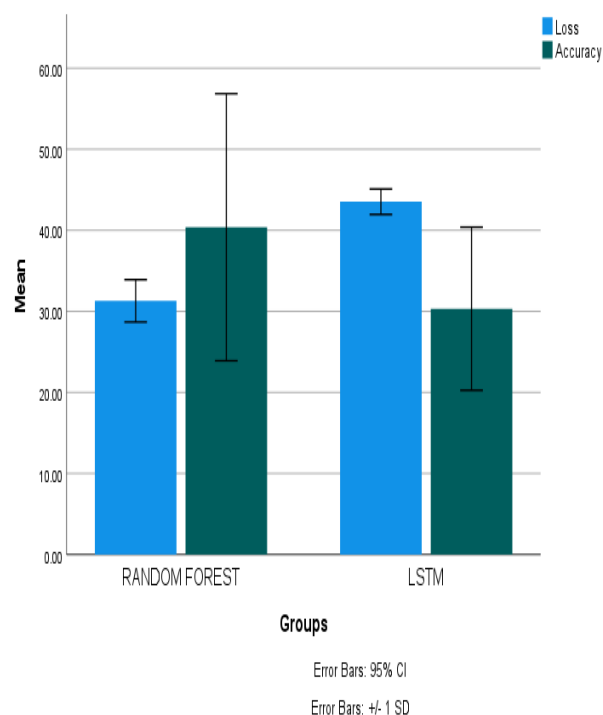


Fig. 1. Comparison of mean accuracy and loss of both randomforest and LSTM.the standard error appears to be less in randomforest compared to LSTM also the standard error appears +/-2SD. X-axis: random forest vs LSTM algorithm. Y-axis : mean accuracy.