International Journal of Modern Science and Technology, 2021;6(5-6):89-94.
ISSN: 2456-0235.
www.ijmst.co

Research Article

Fake news detection using machine learning

S. P. Sivananthan, S. T. Saravanan*, M. Udhai Ram

Department of Computer Science and Engineering,
R. M. K. College of Engineering and Technology, Chennai, India.

*Corresponding author's e-mail: sara17cs080@rmkcet.ac.in

Abstract

The advent of the World Wide Web and the rapid adoption of social media platforms (such as Facebook and Twitter) paved the way for information dissemination that has never been witnessed in the human history before. With the current usage of social media platforms, consumers are creating and sharing more information than ever before, some of which are misleading with no relevance to reality. Automated classification of a text article as misinformation or disinformation is a challenging task. Even an expert in a particular domain must explore multiple aspects before giving a verdict on the truthfulness of an article. In this work, we propose to use machine learning ensemble approach for automated classification of news articles. Our study explores different textual properties that can be used to distinguish fake contents from real. By using those properties, we train a combination of different machine learning algorithms using various ensemble methods and evaluate their performance on 4 real world datasets. Experimental evaluation confirms the superior performance of our proposed ensemble learner approach in comparison to individual learners.

Keywords: Stochastic gradient descent; Term frequency-inverse document frequency; Linear support vector machine; Fake News.

Introduction

The advent of the World Wide Web and the rapid adoption of social media platforms (such as Face book and Twitter) paved the way for information dissemination that has never been witnessed in the human history before. Besides other use cases, news outlets benefitted from the widespread use of social media platforms by providing updated news in near real time to its subscribers. The news media evolved from newspapers, tabloids, and magazines to a digital form such as online news platforms, blogs, social media feeds, and other digital media formats [1].

It became easier for consumers to acquire the latest news at their fingertips. Facebook referrals account for 70% of traffic to news websites [2]. These social media platforms in their current state are extremely powerful and useful for their ability to allow users to discuss and share ideas and debate over issues such as democracy, education, and health. One recent case is the spread of novel corona virus, where fake reports spread over the Internet about the origin, nature, and behavior of the virus [3-9]. The situation worsened as more people read about the fake contents online. Identifying such news online is a daunting task. A more hybrid approach can also be used to analyze the social response of an article along with exploring the textual features to examine whether an article is deceptive in nature or not. A number of studies have primarily focused on detection and classification of fake news on social media platforms such as Facebook and Twitter [13,14].

At conceptual level, fake news has been classified into different types; the knowledge is then expanded to generalize machine learning (ML) models for multiple domains [10,15,16]. The study by Ahmed et al. [17] included extracting linguistic features such as n-grams from textual articles and training multiple ML models including K-nearest neighbor (KNN), support vector machine (SVM), logistic regression (LR), linear support vector machine (LSVM), decision tree (DT), and stochastic gradient descent (SGD), achieving the highest accuracy (92%) with SVM and logistic regression. According to the research, as the number of n increased in n-grams calculated for a particular article, the overall accuracy

Received: 05.05.2021; Received after Revision: 15.05.2021; Accepted: 18.05.2021; Published: 28.06.2021.

decreased. The phenomenon has been observed for learning models that are used for classification.

A convolutional layer is used to capture the dependency between the metadata vectors, followed by a bidirectional LSTM layer. The maxpooled text representations were concatenated with the metadata representation from the bidirectional LSTM, which was fed to fully connected layer with a softmax activation function to generate the final prediction. The research is conducted on a dataset from political domain which contains statements from two different parties. Along with that, some metadata such as subject, speaker, job, state, party, context, and history are also included as a feature set. Accuracy of 27.7% was achieved with combination of features such as text and speaker, whereas 27.4% accuracy was achieved by combining all the different metadata elements with text.

A competitive solution is provided by Riedel et al. [19], which is a stance detection system that assigns one of four labels to an article, “agree,” “disagree,” “discuss,” or “unrelated,” depending on the conformity of article headline with article text. The authors used linguistic properties of text such as term frequency (TF) and term frequency-inverse document frequency (TF-IDF) as a feature set, and a multilayer perceptron (MLP) classifier is used with one hidden layer and a softmax function on the output of the final layer. The dataset contained articles with a headline, body, and label. The system’s accuracy on the “disagree” label on test examples was poor, whereas it performs best with respect to the “agree” label. The authors used a simple MLP with some fine-tuned hyperparameters to achieve an overall accuracy of 88.46%. Shu et al. [12] also discussed several varieties of veracity assessment methods to detect fake news online. Two major categories of assessment methods are explored: one is linguistic cue approaches and the other is network analyses approaches. A combination of both creates a more robust hybrid approach for fake news detection online.

Linguistic approaches involve deep syntax, rhetorical structure, and discourse analysis. These linguistic approaches are used to train classifiers such as SVM or na¨ıve Bayes models. Network-based approaches included analyzing and processing social network behavior and linked data. A unique approach is followed by Vosoughi et al. [13] to explore the properties of news spread on social media; i.e., the authors discussed the spread of news (rumors) on social media such as Twitter and analyzed how the spread of fake news differs from real news in terms of its diffusion on Twitter. Multiple analysis techniques are discussed in the paper to explore the spread of fake news online, such as the depth, the size, the maximum breadth, the structural virality, the mean breadth of true and false rumor cascades at various depths, the number of unique Twitter users reached at any depth, and the number of minutes it takes for true and false rumor cascades to reach depth and number of Twitter users. 1.1. Our Contributions. In the current fake news corpus, there have been multiple instances where both supervised and unsupervised learning algorithms are used to classify text [20,21]. However, most of the literature focuses on specific datasets or domains, most prominently the politics domain [10,19,21]. Therefore, the algorithm trained worksbest on a particular type of article’s domain and does not achieve optimal results when exposed to articles from other domains. Since articles from different domains have a unique textual structure, it is difficult to train a generic algorithm that works best on all news domains. Inthis paper, we propose a solution to the fake news detection problem using the machine learning ensemble approach. Our study explores different textual properties that could be used to distinguish fake contents from real. By using those properties, we train the model using various machine learning methods and arrived to the conclusion that Passive Aggressive classifier is the perfect model for our venture.

Materials and methods

In the following, we describe our proposed framework, followed by the description of algorithms, datasets, and performance evaluation metrics.

Proposed framework

In our proposed framework, as illustrated in Fig. 1, we are expanding on the current literature by introducing ensemble techniques with various linguistic feature sets to classify news articles from multiple domains as true or fake. The ensemble techniques along with Linguistic

Inquiry and Word Count (LIWC) feature set used in this research are the novelty of our proposed approach. There are numerous reputed websites that post legiti mate news contents, and few other websites such as the PolitiFact and Snopes which are used for fact checking. In addition, there are open repositories which are maintained by researchers [11] to keep an up-to-date list of currently available datasets and hyperlinks to potential fact checking sites that may help in countering false news spread. However, we selected three datasets for our experiments which contain news from multiple domains (such as politics, entertainment, technology, and sports) and contain a mix of both truth and fake news. We took our data set from Kaggle, which is available openly for all users. The corpus collected from the World Wide Web is preprocessed before being used as an input for training the models. The articles' unwanted variables such as authors, date posted, URL, and category are filtered out. Articles with no body text or having less than 20 words in the article body are also removed. Multicolumn articles are transformed into single column articles for uniformity of format and Once the relevant attributes are selected after the data cleaning and exploration phase, the next step involves extraction of the linguistic features. Linguistic features involved certain textual characteristics converted into a numerical form such that they can be used as an input for the training models. These features include percentage of wordsimplying positive or negative emotions; percentage of stop words; punctuation; function words; informal language; and percentage of certain grammar used in sentences such as adjectives, preposition, and verbs.

To accomplish the ex- traction of features from the corpus, we used the LIWC2015 tool which classifies the text into different discrete and continuous variables, some of which are mentioned above. LIWC tool extracts 93 different features from any given text. As all of the features extracted using the tool are numerical values, no encoding is required for categorical variables. However, scaling is employed to ensure that various feature's values lie in the range of (0, 1). This is necessary as some values are in the range of 0 to 100 (such as percentage values), whereas other values have arbitrary range (such as word counts). The input features are then used to train the different machine learning models. Each dataset is divided into training and testing set with a 70/30 split, respectively. The articles are shuffled to ensure a fair allocation of fake and true articles in training and tests instances.

Proposed model

The baselines described in namely multi-class classification donevia logistic regression support vector machines. was runThe features used were n-grams and TF-IDF. N-grams are consecutive groups of words, up to the size "n". For example, bi-grams are pairs of words seen next to each other. Features for a sentence or phrase are created from n-grams by having vector that is thr length of the new "vocabulary set" i.e. it has a spot for each unique n-gram that receives a 0 or 1 based on whether n-gram is present in the sentence or not. TF-IDF stands for term frequency inverse document frequency.

It is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. As a feature, TF-IDF can be used for stop- word filtering, i.e. discounting the value of words like "and,", "the", etc. whose counts likely have no effect on the classification of the text. An alternative approach is removing stop-words (as defined in various packages, such as Pythons NLTK). Additionally, we explored some of the characteristic n-grams that may influence Logistic Regression and other classifiers. In calculating the most frequent n-grams for "pant -fire" phrases and those of "true" phrases, we found that the word "wants" more frequently appears in "pants-fire" (i.e. fake news) phrases and the phrase "states" more frequently appears in " true " (i.e. real news)phrases. Intuitively, This makes sense because it's easy to lie about what a politician wants than to lie about what he or she has stated since the former is more difficult to confirm.

Results and discussion

Topic dependency

We took some words that were more common in real news, some that were more common in fake news, and some that were similarly common in both real and fake news type. Fig. 2 shows the distribution of each word in the fake and real news datasets. Also, note that other forms of the word were also included such as plurality. The accuracy in Fig. 3 show how well a model performed on the test set including only articles that contained the given word, after being trained on a dataset that only included articles that did not contain the given word. 3.2 Cleaning. Although pre-processing our data to rid it of any distracting features was an iterative process, we have split it up into three major steps.

These incremental steps each have corresponding models that were trained and tested on the data that was pre-processed at the level represented by the step name. All of the steps build on each other, such that the second step includes the first steps pre-processing and the third step includes the first two pre-processing methods. The first step is simple pre-processing (i.e. tokenization cleaning of data from cationic with the addition of our removal of source, author, title, and date from our own cleaning). Figure 4 shows how the distribution of weights changed as the text was cleaned more. We anticipated that as we removed the easy words which were like cheat codes for classifying the text, there would be more neurons that contributed to the decision of classification and this was confirmed by the standard deviations. The final output of a fully connected layer is computed by summing with ai for each neuron over all neurons.

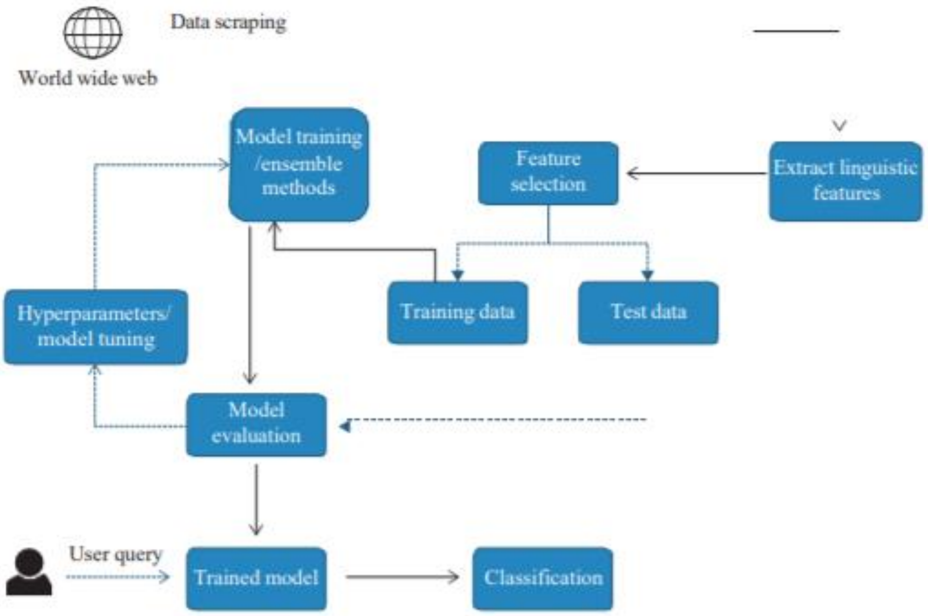
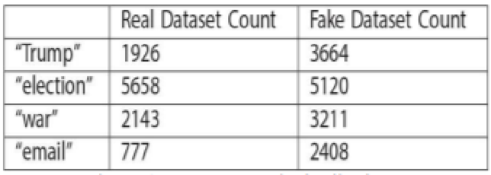


Fig. 1. workflow for training algorithms and classification of news articles

Therefore, the higher the absolute value of ai of a particular neuron, the more importance it had in the final classification decision. Fig. 4 shows how the accuracies of the model changed with more cleaning. We describe how this relates to the standard deviations and vocab size, as seen in Fig. 5. Accuracy of the model changed with cleaning is shown in Fig. 6. The accuracy confusion matrix is shown in Fig. 7. The Fig. 8 shows the train data prediction pie chart.



	Real Dataset Count	Fake Dataset Count
"Trump"	1926	3664
"election"	5658	5120
"war"	2143	3211
"email"	777	2408

Fig. 2. Target word distribution

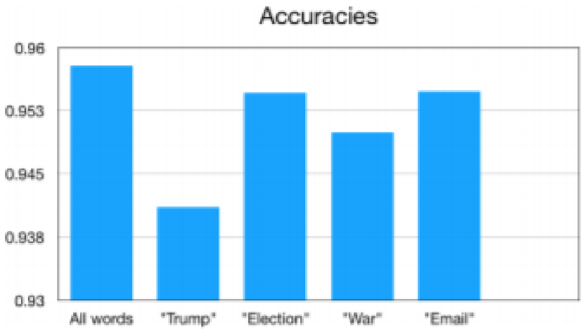


Fig. 3. Accuracies of evaluation using articles with each topic word

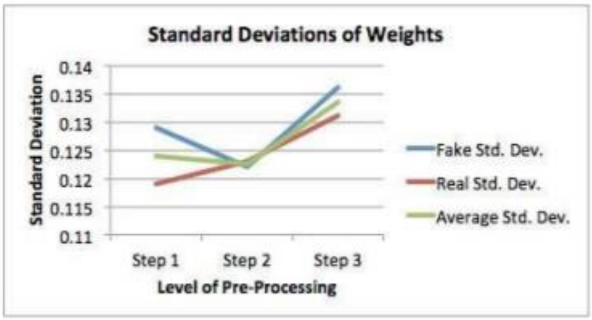


Fig. 4. Standard deviation of neuron weights with cleaning

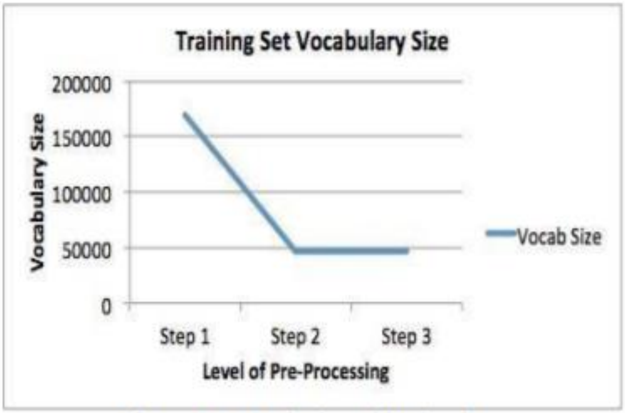


Fig. 5. Vocab size with cleaning

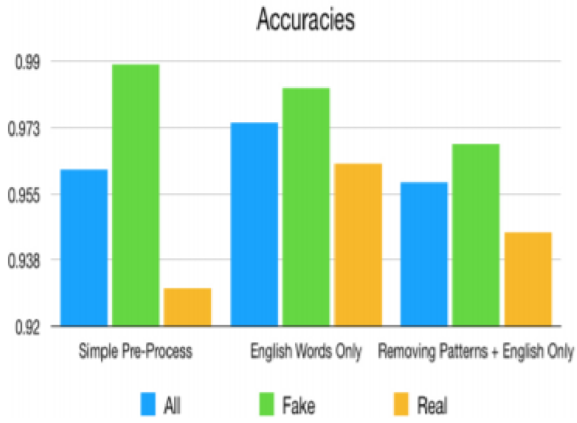


Fig. 6. Accuracies with cleaning

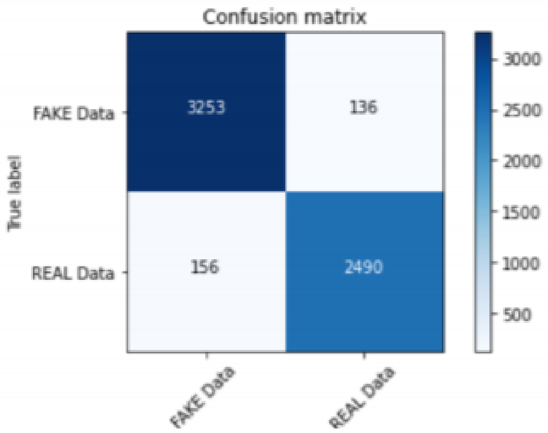


Fig. 7. Accuracy confusion matrix

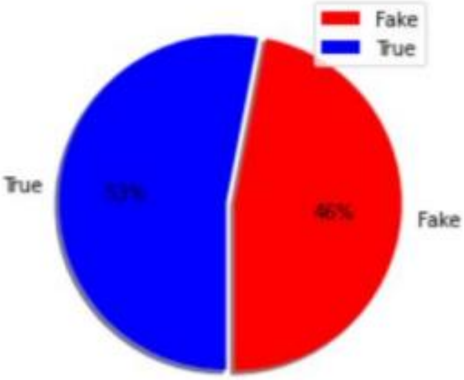


Fig. 8. Train data prediction pie chart

Conclusions

In this work, authors have analyzed the process of data collection, data cleaning, data processing and the training of model by considering a number of models for execution. Using pictorial analysis and accuracy analysis we have picked Passive aggressive classifier for our analysis. This model proved to be highly accurate for our data set that we collected from Kaggle. We successfully trained the model and provided the result as graph for better understanding. In addition to that we have implemented the model on a test dataset and provided the output as pie chart for better understanding. Finally, we calculated the accuracy for both fake and real news and including as an overlay in the pie chart.

Conflict of interest

Authors declare no conflict of interest.

References

[1] Douglas A. News consumption and the new electronic media. The International Journal of Press/Politics. 2006;11:29-52.

[2] Wong J. Almost all the traffic to fake news sites is from facebook. New data show, 2016.

[3] Lazer DMJ, Baum MA, Benkler Y, et al., The science of fake news. Science 2018;3591094–1096.

[4] Garc´ıa SA, Garc´ıa GG, Prieto MS, Guerrero AJM, Jime´nez CR. The impact of term fake news on the scientific community scientific performance and mapping in web of science. Social Sciences 2020;9:73. https://doi.org/10.3390/socsci9050073

[5] Holan AD. Lie of the Year: Fake News, Politifact, Washington, DC, USA, 2016.

[6] Kogan S, Moskowitz TJ, Niessner M. Fake News: Evidence from Financial Markets, 2019. https://ssrn.com/ abstract=3237763.

[7] Robb A. Anatomy of a fake news scandal," Rolling Stone. 2017;1301:28–33.

[8] Soll J. The long and brutal history of fake news. Politico Magazine 2016;18.

[9] Hua J, Shaw R. Corona virus (covid-19) "infodemic" and emerging issues through a data lens: the case of China. International Journal of Environmental Research and Public Health 2020;17:2309.

[10] Conroy NK, Rubin VL, Chen Y. Automatic deception detection: methods for finding fake news. Proceedings of the Association for Information Science and Technology 2015;52:1–4.

[11] Asr FT, Taboada M. Misinfotext: a collection of news articles, with false and true labels, 2019.

[12] Shu K, Sliva A, Wang S, Tang J, Liu H. Fake news detection on social media. ACM SIGKDD Explorations Newsletter 2017;19:22–36.

[13] Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018;359:1146–1151.

[14] Allcott H, Gentzkow M. Social media and fake news inthe 2016 election. Journal of Economic Perspectives 2017;31:211–236.

[15] Rubin VL, Conroy N, Chen Y, Cornwell S. Fake news or truth? using satirical cues to detect potentially misleading news," in Proceedings of the Second Workshop on Computa- tional Approaches to Deception Detection, San Diego,CA, USA, 2016, pp. 7–17.

[16] Jwa H, Oh D, Park K, Kang JM, Lim H. exBAKE: automatic fake news detection model based on bidirectional encoder representations from transformers (bert). Applied Sciences 2019;9:4062. https://doi.org/10.3390/app9194062

[17] Ahmed H, Traore I, Saad S. Detection of online fake news using n-gram analysis and machine learning techniques. Proceedings of the International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, pp. 127–138, Springer, Vancouver, Canada, 2017.

[18] Wang WY. Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection, Association for Computational Linguistics, Stroudsburg, PA, USA, 2017.

[19] Sriram A, Sudhakar TD. Technology revolution in the inspection of power transmission lines - A literature review. 7th International Conference on Electrical Energy Systems (ICEES), 2021, pp. 256-262.
doi: 10.1109/ICEES51510.2021.9383707.

[20] Anbalagan S, Sudhakar TD. Protection of Power Transmission Lines Using Intelligent Hot Spot Detection. Fifth International Conference on Electrical Energy Systems (ICEES), 2019, pp. 1-6.
doi: 10.1109/ICEES.2019.8719290

[21] Riedel B, Augenstein I, Spithourakis GP, Riedel S. A simple but tough-to-beat baseline for the fake news challenge stance detection task. 2017. https://arxiv.org/abs/1707. 03264.

****