IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11541

ANALYSIS OF FAKE NEWS DETETCTION USING MACHINE LEARNING TECHNIQUES

Monal Eswar. N, Padmapriya. V, Prabavathi. V, Lilly Florence. M

CSE, Adhiyamaan College of Engineering (Autonomous), Anna University, Hosur, India

Abstract: News online has become the major source of information for people, much information appearing on the internet is dubious and even intended to mislead. Automated fake news detection tools like machine learning and deep learning models have become an essential requirement also used stemming, lemmatization, stop word techniques to obtain text representation for machine learning and deep learning models respectively. We use Kaggle dataset, for defining the fake news. This would allow to provide a filtered subset of fake news to end users. 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 has to 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. Along with the data, our understanding of AI also increases and the computing power enables us to train very complex and large models faster. Fake news has been gathering a lot of attention worldwide recently. The effects can be political, economic, organizational, or even personal. This paper discusses the approach of natural language processing and machine learning in order to solve this problem. Use of bag-of-words, n-grams, count vectorizer has been made, TF-IDF, and trained the data on five classifiers to investigate which of them works well for this specific dataset of labelled news statements. The precision, recall and f1 scores help us determine which model works best.

Keywords: Fake news analysis, real news, Keywords Internet, social media, Fake News, Classification, Machine Learning.

1.INTRODUCTION

As time flows, the amount of data, especially text data increases exponentially. Along with the data, our understanding of AI also increases and the computing power enables us to train very complex and large models faster. Fake news has been gathering a lot of attention worldwide recently. The effects can be political, economic, organizational, or even personal. This paper discusses the approach of natural language processing and machine learning in order to solve this problem. Use of bag-of-words, n-grams, count vectorizer has been made, TF-IDF, and trained the data on five classifiers to investigate which of them works well for this specific dataset of labelled news statements. The precision, recall and f1 scores help us determine which model works best. 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. 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. It became easier for consumers to acquire the latest news at their fingertips. Facebook referrals account for 70% of traffic to news websites. 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. However, such platforms are also used with a negative perspective by certain entities commonly for monetary gain [3, 4] and in other cases for creating biased opinions, manipulating mindsets, and spreading satire or absurdity. The phenomenon is commonly known as fake news.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11541

LITERATURE SURVEY

2. RELATED WORK

The majority of previous studies have focused on categorising internet news and social media articles. Various investigations have offered various strategies for detecting deceit. Fake news may be divided into several categories. Conroy, Rubin, and Chen, for example, have identified three forms of false news: Serious Forgeries (Type A), Large-Scale Hoaxes (Type B), and Humorous Forgeries (Type C) [20]. In simple terms, fake news is a news piece that is purposefully and verifiably untrue and may cause readers to be misled [2]. This specific definition is advantageous in that it distinguishes between false news and other similar ideas such as hoaxes and satires. 2 https://www.bbc.co.uk/news/resources/idt-sh/nigeria fake news 3 https://anonymous.4open.science/repository/b7c0d56e-9e4b-434b-87f4-16d9a0f0516/ 2 Linguistic-based parameters such as total words, characters per word, frequency of large words, frequencies of phrases (i.e., n-grams and bag-of-words techniques [9]), and parts-of-speech (POS) tagging have been proposed by Shu, Silva, Wang, Jiliang, and Liu [22]. Simple content-related n-grams and part-of-speech (POS) tagging has been shown to be insufficient for the classification job, according to Conroy, Rubin, and Chen [6]. Instead, they recommended Deep Syntax Analysis with Probabilistic Context-Free Grammars (PCFG), citing Feng, Banerjee, and Choi [8] who utilised this technique to discriminate rule types (lexicalized, non-lexicalized, parent nodes, and so on) for deception detection. However, whereas bi-gram TF-IDF provides extremely successful models for identifying false news, the PCFG characteristics add nothing to the model's efficacy [10], according to Shlok Gilda. Many studies have proposed using sentiment analysis to identify deceit since there may be a link between the sentiment of a news story and its nature. Conroy, Rubin, Chen, and Cornwell believed that quantifying the usefulness of characteristics like part of speech frequency and semantic categories such generalising terms, positive and negative polarity (sentiment analysis) would broaden the possibilities of word-level analysis [19]. In his sarcasm detection blog, Mathieu Cliche describes how n-grams, terms acquired from tweets expressly marked as sardonic, may be used to identify sarcasm on Twitter. To increase prediction accuracy, he also uses sentiment analysis and subject identification (words that are frequently clustered together in tweets) [5]. On their suggested dataset LIAR [26], Wang compared the performance of SVM, LR, Bi-LSTM, and CNN models. Several studies have shown good results in identifying bogus news and tracing user propagation using neural networks. Wang developed a hybrid convolutional neural network model that outperforms other standard machine learning algorithms in his [26]. Hannah Rashkin et al. [18] conducted a thorough investigation of linguistic variables and demonstrated the power of LSTM. Singhania et al. [23] presented a three-level hierarchical attention network with one level for words, phrases, and a news article's title. Ruchansky et al. [21] developed the CSI model, which captures content, an article's reaction, and source attributes based on user behaviour.

3. METHODOLOGY

This paper comes up with the applications of NLP (Natural Language Processing) techniques for detecting the 'fake news', that is misleading news stories that comes from the non-reputable sources. In this paper a model is built based on the count vectorizer or a tf-idf matrix (i.e. word tallies relatives to how often they are used in other articles in your dataset) can help. Since this problem is a kind of text classification, implementing a Naive Bayes classifier and Logistic Regression will be better as this is standard for text-based processing. The actual goal is in developing a model which was the text transformation (count vectorizer vs tf-idf vectorizer) and choosing which type of text to use (headlines vs full text). Now the next step is to extract the most optimal features for count vectorizer or tfidf-vectorizer, this is done by using a n-number of the most used words, and/or phrases, lower casing or not, mainly removing the stop words which are common words such as "the", "when", and "there" and only using those words that appear at least a given number of times in a given text dataset.

The performance of a classifier may vary based on the size and quality of the text data and also the features of the text vectors. Common noisy words called 'stopwords' are less important words when it comes to text feature extraction, they don't contribute towards the actual meaning of a sentence and they only contribute towards feature dimensionality and may be discarded for better performance. This helps in reducing the size/dimensionality of the text corpus and add text context for feature extraction.

Lemmatization is used to convert words to their core meaning and this results in multiple word conversion into a single discrete representation. Finally, the analysis process will take place and concludes which algorithm performance is best.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11541

Figure 1. Architecture of fake news detection

3.1 Algorithms

We used the following learning algorithms in conjunction with our proposed methodology to evaluate the performance of fake news detection classifiers.

3.1.1 Logistic Regression

A logistic regression (LR) model is used to classify text based on a large feature set with a binary output (true/false or real article/fake article), since it gives a straightforward equation to classify issues into binary or many classes [27]. We tuned hyperparameters to acquire the best results for each dataset, and we evaluated numerous values before getting the greatest accuracies from the LR model. The logistic regression hypothesis function can be described mathematically as follows [27]:

$$h_\theta(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}.$$

Logistic regression uses a sigmoid function to transform the output to a probability value; the objective is to minimize the cost function to achieve an optimal probability. The cost function is calculated as shown in

$$\text{Cost}\left(h_\theta(x), y\right) = \begin{cases} \log\left(h_\theta(x)\right), & y = 1, \\ -\log\left(1 - h_\theta(x)\right), & y = 0. \end{cases}$$

3.1.2 Support Vector Machine

Another model for binary classification problems is the support vector machine (SVM), which is available in a variety of kernel functions [28]. An SVM model's goal is to categorise data points by estimating a hyperplane (or decision boundary) based on a feature set [29]. The size of the hyperplane is determined by the number of features. Because a hyperplane might exist in several places in an N-dimensional space, the goal is to find the plane that separates the data points of two classes with the greatest margin. The cost function for the SVM model is mathematically represented as described in [30] and illustrated in such a way that

$$J(\theta) = \frac{1}{2}\sum_{j=1}^{n}\theta_j^2,$$

such that

$$\theta^T x^{(i)} \geq 1, \quad y^{(i)} = 1,$$
$$\theta^T x^{(i)} \leq -1, \quad y^{(i)} = 0.$$

The function above uses a linear kernel. Kernels are usually used to fit data points that cannot be easily separable or data points that are multidimensional. In our case, we have used sigmoid SVM, kernel SVM (polynomial SVM), Gaussian SVM, and basic linear SVM models.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11541

3.1.3 K-Nearest Neighbours (KNN)

KNN is an unsupervised machine learning model that does not require the use of a dependant variable to predict the result of a given set of data. We give the model enough training data and let it pick which neighbourhood a data point belongs to. The KNN model calculates the distance between a new data point and its closest neighbours, and the value of K calculates the majority of its neighbours' votes; if K is 1, the new data point is allocated to the class with the shortest distance. The following are the mathematical formulas for calculating the distance between two places [31]:

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2},$$

$$\text{Manhattan distance} = \sum_{i=1}^{k}|x_i - y_i|,$$

$$\text{Minkowski distance} = \left(\sum_{i=1}^{k}|x_i - y_i|^q\right)^{1/q}.$$

3.1.4 Naïve Bayes Classifier Algorithm

The supervised learning technique known as the Nave Bayes algorithm is based on the Bayes theorem and is used to solve classification issues. It is mostly utilised in text classification tasks that need a large training dataset.

The Nave Bayes Classifier is a simple and effective classification method that aids in the development of rapid machine learning models capable of making quick predictions.

It's a probabilistic classifier, which means it makes predictions based on an object's likelihood. Spam filtration, sentiment analysis, and article classification are all common uses of the Nave Bayes Algorithm.

3.2 Performance Metrics

We employed a variety of indicators to assess algorithm performance. The confusion matrix is used in the majority of them. The confusion matrix is a tabular representation of the performance of a classification model on the test set, with four parameters: true positive, false positive, true negative, and false negative.

3.2.1 Accuracy

Accuracy is a commonly used measure that represents the percentage of accurately anticipated true or erroneous observations. The following equation can be used to calculate the accuracy of a model's performance:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}.$$

In most cases, high accuracy value represents a good model, but considering the fact that we are training a classification model in our case, an article that was predicted as true while it was actually false (false positive) can have negative consequences; similarly, if an article was predicted as false while it contained factual data, this can create trust issues. Therefore, we have used three other metrics that take into account the incorrectly classified observation, i.e., precision, recall, and F1-score.

3.2.2 Recall

Recall represents the total number of positive classifications out of true class. In our case, it represents the number of articles predicted as true out of the total number of true articles.

$$\text{Recall} = \frac{TP}{TP + FN}.$$

3.3.3 Precision

Precision score, on the other hand, is the ratio of true positives to all real events anticipated. Precision in this example refers to the number of articles tagged as true out of all the positively predicted (true) articles:

$$\text{Precision} = \frac{TP}{TP + FP}.$$

3.3.4 AUC

The name AUC stands for “area under the curve”. The curve in our case is the receiver operating characteristics (ROC) curve. It’s a statistical measure that we can use to evaluate the model predictions using a probabilistic framework.Briefly,

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11541

the ROC curve shows the relationship between false-positive rate and true positive rate for different probability thresholds of model predictions.

3.3 Dataset

George McIntire created the Fake or Real News dataset6. The fake news element of this dataset was sourced from the Kaggle fake news dataset, which included material from the 2016 US presidential election. For the course of 2015 or 2016, genuine news was gathered from media outlets such as the New York Times, WSJ, Bloomberg, NPR, and the Guardian. The dataset's GitHub repository has around 7.8k news items, with an equal mix of false and real news, with half of the corpus coming from political news.

3.3.1 Pre-processing data

The bulk of social media data is unstructured communication, which includes typos, slang, and poor language. In the search for better performance and dependability, the necessity to create methods for utilising resources to make informed decisions has become crucial. The data must first be cleansed before predictive modelling may be used to get further insights.

3.3.2 Vectorizing Data

Vectorizing is the process of converting text into integers (numbers) so that machine learning algorithms can interpret it.

- Bag-Of-Words
- N-Grams
- TF-IDF

a. Vectorizing Data: Bag-Of-Words

The presence of words in text data is described by the Bag of Words (BoW) or Count Vectorizer. It returns 1 if the word is present in the phrase and 0 if it is not. As a result, each text document generates a bag of words with a document-matrix count.

b. Vectorizing Data: N-Grams

In our provided text, n-grams are just all conceivable combinations of neighbouring words or letters of length n. Unigrams are ngrams with n=1 as the number of letters. The same rules apply to bigrams (n=2), trigrams (n=3), and so on. Unigrams are often short and concise.

In contrast to bigrams and trigrams, the fundamental concept of n-grams is that they capture the letter or word that is most likely to follow the given word. The longer the n-gram, the more background you have to deal with (greater n).

c. Vectorizing Data: TF-IDF

It calculates the "relative frequency" of a term in a document compared to the frequency of that word across all papers. The relative importance of a phrase in the document and throughout the corpus is represented by TF-IDF weight.

Term Frequency is abbreviated as TF. It determines how many times a term appears in a document. Because document sizes vary, a term may appear more frequently in a long document than in a small one. As a result, the document's length frequently divides Term frequency. Search engine scoring, text summarization, and document clustering all employ this. TF (t, d) =number of times t appears in document 'd' divided by total number of words in document 'd'. Inverse Document Frequency (IDF) stands for If a word appears in every document, it is of little use. Certain phrases, such as "a," "an," "the," "on," "of," and others, appear frequently in a document yet have little meaning. The value of these terms is reduced by IDF, while the importance of rare terms is increased. The greater the value of IDF, the more distinct the word becomes. IDF (t, d) =total number of documents divided by number of documents containing the phrase t The relative count of each word in each sentence is recorded in the document matrix after TF-IDF is applied to the body text.

IDF TFIDF (t, d) =TF (t, d) * (f)

Sparse matrices are produced using vectorizers.

4. RESULT

Accuracy

The accuracy of the passive aggressive classifier is 0.98, the accuracy of logistic regression is 0.97, the accuracy of gaussian NB is 0.94, and the accuracy of the K-Neighbours classifier is 0.57.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11541

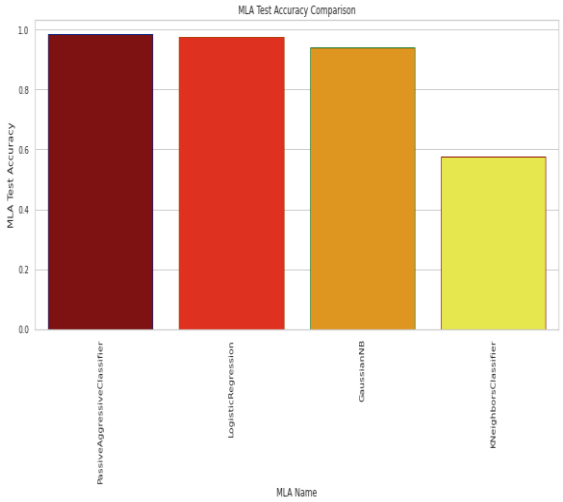

Figure 2. Accuracy diagram for precision comparison

Precision

The precision of the passive aggressive classifier is 0.98, the precision of the logistic regression is 0.96, the precision of the gaussian NB classifier is 0.93, and the precision of the K-Neighbours classifier is 0.94.

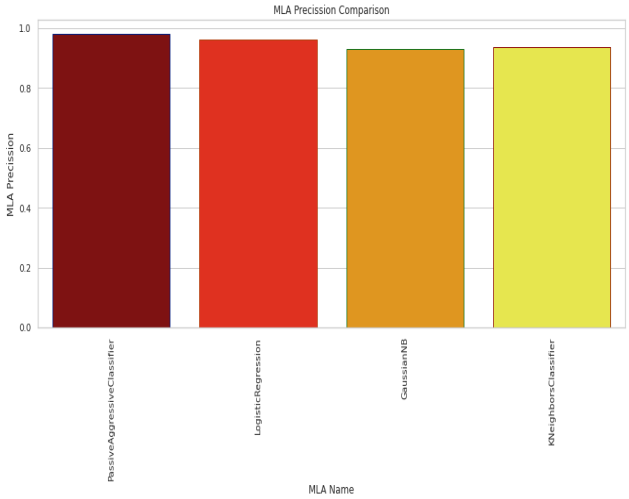

Figure 3. Precision diagram

Recall

The recall of a passive aggressive classifier is 0.99, that of a logistic regression is 0.99, that of a gaussian NB is 0.95, and that of a K-Neighbours classifier is 0.15.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11541

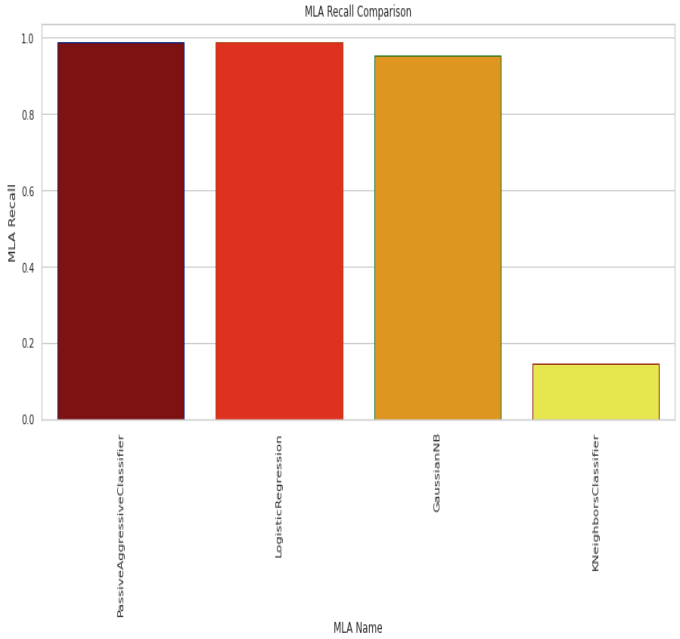

Figure 4. Recall diagram

AUC

The AUC of the passive aggressive classifier is 0.98, the AUC of logistic regression is 0.97, the AUC of gaussianNB is 0.94, and the AUC of the KNeighbours classifier is 0.57.

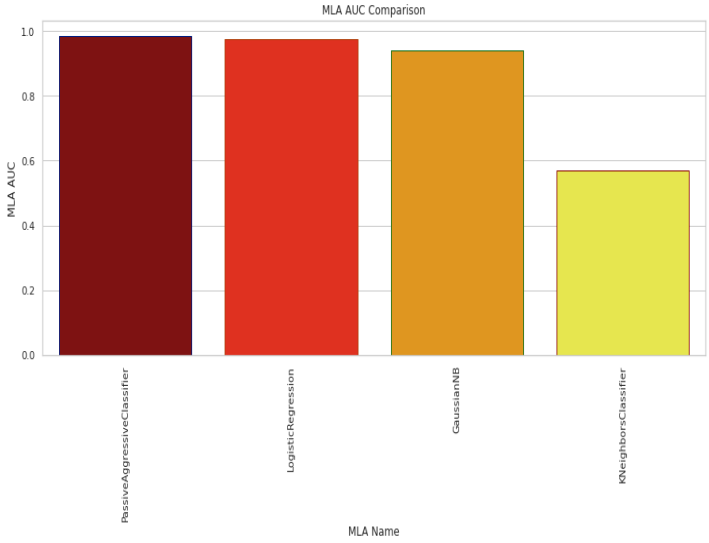

Figure 5. AUC diagram

5. CONCLUSION

The task of classifying news manually requires in-depth knowledge of the domain and expertise to identify anomalies in the text. In this project, we discussed the problem of classifying fake news articles using machine learning models. The dataset we used in our work is collected from the kaggle and contains news articles from various domains to cover most of the news rather than specifically classifying political news.

The primary aim of the research is to identify patterns in text that differentiate fake articles from true news. We used four algorithms for analysis and found which algorithm is best and time consuming . The learning models were trained and parameter-tuned to obtain optimal accuracy. Some models have achieved comparatively higher accuracy than others. We used multiple performance metrics to compare the results for each algorithm. The ensemble learners have shown an overall better score on all performance metrics as compared to the individual learners.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11541

REFERENCES

[1] Hadeer Ahmed, Issa Traore, and Sheriff Saad. Detection of online fake news using n-gram analysis and machine learning techniques. In International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, pages 127–138. Springer,2017.

[2] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2):211–36, 2017.

[3] Peter Bourgonje, Julian Moreno Schneider, and Georg Rehm. From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles. In Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism, pages 84–89, 2017.12

[4] Yimin Chen, Niall J Conroy, and Victoria L Rubin. Misleading online content: Recognizing clickbait as false news. In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, pages 15–19. ACM,2015.

[5] Mathieu Cliche. The sarcasm detector, 2014.

[6] Niall J Conroy, Victoria L Rubin, and Yimin Chen. Automatic deception detection: Methods for finding fake news. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, page 82. American Society for Information Science, 2015.

[7] Ethan Fast, Binbin Chen, and Michael S Bernstein. Empath: Understanding topic signals in large-scale text. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pages 4647–4657. ACM, 2016.

[8] Song Feng, Ritwik Banerjee, and Yejin Choi. Syntactic stylometry for deception detection. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2, pages 171–175.Association for Computational Linguistics, 2012.

[9] Johannes Furnkranz. A study using n-gram features for text categorization. ¨ Austrian Research Institute for Artifical Intelligence, 3(1998):1–10, 1998.

[10] Shlok Gilda. Evaluating machine learning algorithms for fake news detection. In Research and Development (SCOReD), 2017 IEEE 15th Student Conference on, pages 110–115. IEEE, 2017.

[11] Mykhailo Granik and Volodymyr Mesyura. Fake news detection using naive bayes classifier. In Electrical and Computer Engineering (UKRCON), 2017 IEEE First Ukraine Conference on, pages 900–903. IEEE, 2017.

[12] Angel Hern ´ andez-Casta ´ neda and Hiram Calvo. Deceptive text detection using continuous semantic space models. ˜ Intelligent Data Analysis, 21(3):679–695, 2017.

[13] Johan Hovold. Naive bayes spam filtering using word-position-based attributes. In CEAS, pages 41–48, 2005.

[14] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759, 2016.

[15] David Leonhardt and Stuart A Thompson. Trumps lies. New York Times, 21, 2017.

[16] Rada Mihalcea and Carlo Strapparava. The lie detector: Explorations in the automatic recognition of deceptive language. In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, pages 309–312. Association for Computational Linguistics, 2009.

[17] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pages 1532–1543, 2014.

[18] Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. Truth of varying shades: Analysing language in fake news and political fact-checking. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 2931–2937, 2017.

[19] Victoria Rubin, Niall Conroy, Yimin Chen, and Sarah Cornwell. Fake news or truth? using satirical cues to detect potentially misleading news. In Proceedings of the Second Workshop on Computational Approaches to Deception Detection, pages 7–17, 2016.

[20] Victoria L Rubin, Yimin Chen, and Niall J Conroy. Deception detection for news: three types of fakes. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, page 83. American Society for Information Science, 2015.

[21] Natali Ruchansky, Sungyong Seo, and Yan Liu. Csi: A hybrid deep model for fake news detection. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pages 797–806. ACM, 2017.13

[22] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1):22–36, 2017.

[23] Sneha Singhania, Nigel Fernandez, and Shrisha Rao. 3han: A deep neural network for fake news detection. In International Conference on Neural Information Processing, pages 572–581. Springer, 2017.

[24] Eugenio Tacchini, Gabriele Ballarin, Marco L Della Vedova, Stefano Moret, and Luca de Alfaro. Some like it hoax: Automated fake news detection in social networks. arXiv preprint arXiv:1704.07506, 2017.

[25] James Thorne, Mingjie Chen, Giorgos Myrianthous, Jiashu Pu, Xiaoxuan Wang, and Andreas Vlachos. Fake news stance detection using stacked ensemble of classifiers. In Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism, pages 80–83, 2017.

DOI: 10.17148/IJARCCE.2022.11541

[26] William Yang Wang." liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648, 2017.
[27] Xingyou Wang, Weijie Jiang, and Zhiyong Luo. Combination of convolutional and recurrent neural network for sentiment analysis of short texts. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 2428–2437, 2016.
[28] Liang Wu, Jundong Li, Xia Hu, and Huan Liu. Gleaning wisdom from the past: Early detection of emerging rumors in social media. In Proceedings of the 2017 SIAM International Conference on Data Mining, pages 99–107. SIAM, 2017.
[29] Liang Wu and Huan Liu. Tracing fake-news footprints: Characterizing social media messages by how they propagate. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, pages 637–645. ACM, 2018.
[30] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1480–1489, 2016.
[31] Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-level convolutional networks for text classification. In Advances in neural information processing systems, pages 649–657, 2015.
[32] Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis Lau. A c-lstm neural network for text classification. arXiv preprint arXiv:1511.08630, 2015.