Int J Syst Assur Eng Manag
https://doi.org/10.1007/s13198-020-01016-4

ORIGINAL ARTICLE

Study and analysis of unreliable news based on content acquired using ensemble learning (prevalence of fake news on social media)

Mohammad Zubair Khan[1] · Omar Hussain Alhazmi[1]

Received: 29 December 2019 / Revised: 21 June 2020

Abstract We explore the use of machine learning techniques to classify a news source for generating unreliable news. Since the advent of the Internet, unreliable news and hoaxes have deceived users. Social media and news outlets are spreading false information to increase the number of viewers or as a part of the psychological competition. In this paper, we present an ensemble classifier using a set of marked true and bogus news articles. Here, the authors develop a classification approach based on text using SVM, Random-Forest, Naïve Bayes, Decision Tree as a base learner in Bagging and AdaBoost. The purpose behind the work is to think of an answer that enable the user to classify and filter some of the false material. Accordingly, we show that the best performing classifiers were AdaBoost-LinearSVM and AdaBoost-Random Forest with 90.70% and 80.17% accuracy, respectively.

Keywords Ensemble learning · AdaBoost · Bagging · Fake news · Social media

✉ Omar Hussain Alhazmi
ohhazmi@taibahu.edu.sa

Mohammad Zubair Khan
zubair.762001@gmail.com

[1] Department of Computer Science, College of Computer Science and Engineering, Taibah University, Madinah, Saudi Arabia

1 Introduction

Fake news is not an illusion of innovation. Even before the Internet was used as a source, dissemination of misleading and inaccurate information was prevalent. Unreliable news, defined as a made-up story to deceive, has been broadly referred to as a contributing factor to the result of the 2019 and 2014 Indian Parliamentary Elections. While Mark Zuckerberg, Facebook's CEO, admitted that Facebook has unreliable news, Facebook and other online news sources have started to create procedures for distinguishing unreliable from reliable news and mitigating its spread. Zuckerberg (2016) accepted that identifying counterfeit news is difficult, stating, "This is a region where I accept, we should continue cautiously; however, distinguishing the truth is complex." Spicer similarly observed: "Click baits are phrases designed to attract the attention of a user who, after flicking the URL-Link, is coordinated with such a website page whose content is far below their expectations" (Spicer 2018). Several other users find click baits a bother, and the result is that most of these people end up visiting such websites only for a very short time. Consequently, unreliable news is progressively turning into a hazard and a concern to societies and individuals. It is commonly created for business interests to attract users for financial gain. Moreover, It is now widely accepted that individuals and groups with possibly wicked agendas have been known to start fake/unreliable news to influence events and policies around the globe. For content distributors, however, more clicks equates to more income, as the business part of utilising the web advertisements is highly dependent upon web traffic (Conroy et al. 2015). Specifically, many websites containing this information incorporate a sharing option that requires users to further

Published online: 13 July 2020

disseminate the content of the website page. Social media sites allow information to be exchanged efficiently and quickly, while motivating users to exchange unreliable data in a short time. In the aftermath of Cambridge Analytica's data break of millions of records, Facebook and other mammoths vowed to do more to prevent false and inaccurate news spreading (Smith et al. 2018). Therefore, this work employs ensemble learning techniques to detect the fake news; we conduct and experiment with a sizable data set and apply experiment assumptions and settings and then present and discuss the outcoming results. This result can be helpful to social media outlets and news websites and law enforcement agencies. Hence, this will make news more reliable and will help people get the true news and promote honest online news publishing practices. Eventually, leading to more professional and healthy media industry.

The organization of this paper is as follows: in the next section we present literature review of the current studies in the fake news research and other related studies. Then, in Sect. 3, we discuss methodology and model design; the experiment will be presented in Sect. 4. Finally, the results are discussed in Sect. 5 and the conclusion in Sect. 6.

2 Literature Review

Some research argues that unreliable news still has some benefits for brands listed on the stock market that have experienced price growth due to unreliable news (Hassid 2011). As the number of users accessing a certain company's website containing false data about the company's product and services, potential investors are taking an interest in the company's operations. Therefore, this may cause some customers to buy unjustly overpriced shares. Thus, this will lead to regret, as most people who buy shares based on dishonesty end up being disappointed. Likewise, different authors noted that inaccurate news can help further the promotion of a company's marketing goals. For example, Marchi noted that "if the information on the website pages relating to such news is one that supports the products supplied by a company, more customers will be enthusiastic about the equivalent even though the content of the website page is far from the truth" (Marchi 2012). According to Rubin and colleagues, "Nonetheless, such an entity interacts with a wider pool of prospective consumers even though the unreliable news was not part of its marketing efforts. The researchers argue that the concept of unreliable news is not completely terrible as it can positively contribute to an enterprise's growth. Look at contemporary academic work shows that the issue of unreliable news was a major concern among researchers from various backgrounds. Some authors have seen, for example, that unreliable news is never again a marketing and advertising project" (Rubin et al. 2015). This problem is slowly being regarded as a major aspect of the IT division duties. It was generally accepted that the two offices referred to above were responsible for managing any difficulties arising from the dispersal of false news identified with a company. Current research indicates that inaccurate news is seen as a threat to information protection. The IT division's involvement is thus prefaced with the possibility that it will help to avert the various hazards associated with the problem.

The current online platforms and social media create useful tools for spreading false stories broadly. Accordingly, writers and certainty checkers with their present systems cannot classify unreliable stories continuously before they are out of control. Computerising those techniques is one way to accelerate the process. This issue is considered to fit a machine learning task (Markowitz and Hancock 2014; Hardalov et al. 2016). Thus far, the work on fighting counterfeit news used to be handled within many separate activities and studies. However, organisations like FullFact.org recommends initiating joint efforts between these tasks to add a stage that gives a gathering of devices to deal with the different aspects of fact-checking routines2. Correspondingly, Fake News Challenge (FNC-1), which is an online competition, additionally recommends a solution for unreliable news recognition based on accumulation of mechanised tools to allow human fact-checkers work easier. Stance discovery has been demonstrated to be helpful in disinformation recognition. Jin et al. investigated the credibility of spread of news verification through building associations between micro-blog sites (tweets) as supporting or denying others' perspectives (Jin et al. 2016). The veracity of cases was likewise predicted using the position of articles and the unwavering quality of their sources (Popat et al. 2017). Position features were also used in distinguishing the believability of bits of gossip, which are also defined to be unverified claims (Enayet and El-Beltagy 2017). Moreover, using Tweets distributing time and positions as the main features to show the stance of tweets using Hidden Markov Models accomplished high precision (Dungs et al. 2018). Identifying the stance of news stories is most closely related to our task. On this line crafted by Ferreira and Vlachos addresses gossip exposing based on stance. The point is to estimate the position of a news headline towards its matched case as Observing, For or Against. Linguistic features are separated from each claim and headline pair (Ferreira and Vlachos 2016). Moreover, Alduwari and Alwehaidi (2018) have used feature selection to capture fake news and clickbaits; clickbaits are deceptive links that use flashy headlines to lure users into clicking a link to some advertised website (Gardiner 2015). Recently, Ozbay and Alatas have

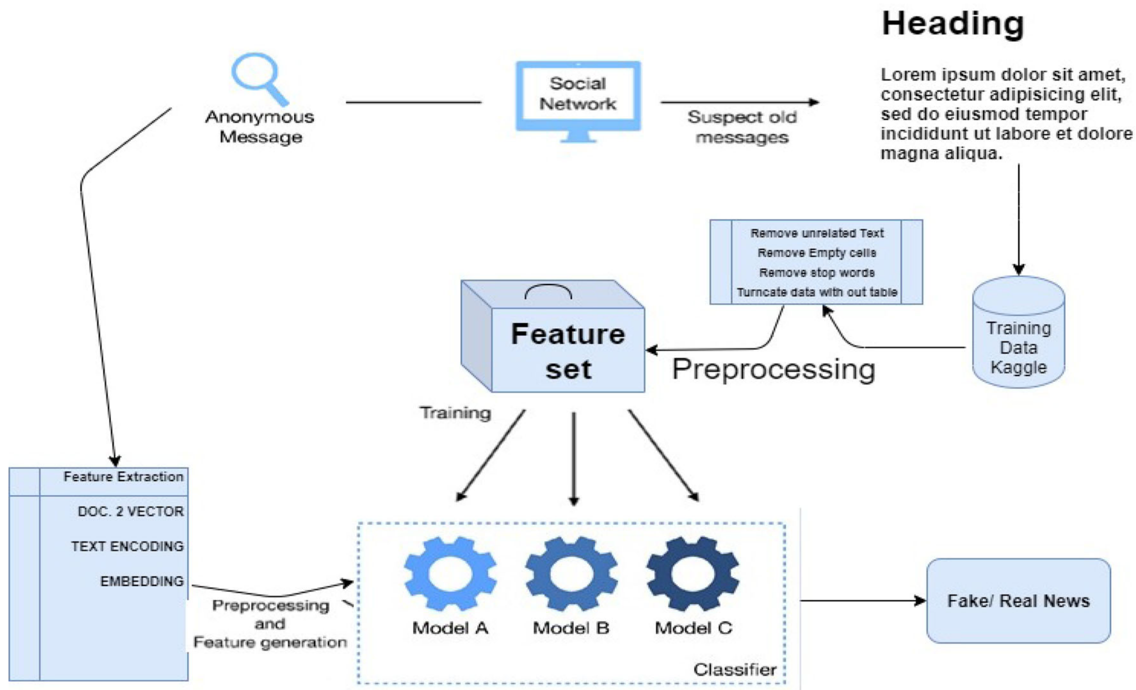

Fig. 1 Workflow of the system

compared twenty-three different artificial intelligence algorithms in order to capture fake news and hoaxes (Ozbay and Alatas 2020). Reis et al. have examined categorized fake news; they have shown strong connection that feature selection can be tailored to target different categories of fake news with promising results.

3 Methodology

The unreliable news issue needs a solution of a model that can identify and eliminate unreliable websites from the results that a browser or news feed to social media provides to a user. The model aims to assist in recognising unreliable news sources in light of various articles related to that source. If a site is classified as an unreliable news provider, we can categorically state that any future articles from that network will be deemed unreliable news.

Table 1 Confusion Matrix

	Predicted: NO	Predicted: YES	Total
Actual: NO	TN	FP	N
Actual: YES	FN	TP	p
Total	N’	P’	

3.1 Dataset

The dataset used for this proposed solution was obtained from Kaggle (Kaggle 2019). In this dataset, the training data comprises of about 20,008 rows of various articles published on the Internet. In this solution first, do preprocess of data to train the proposed model (Fig. 1).

“A training dataset has the following useful attributes:

a. id: unique id for a news article
b. title: the title of a news article
c. author: author of the news article
d. text: the text of the article; incomplete in some cases
e. label: a label that marks the article as potentially unreliable” (Kaggle 2019).

- 1: unreliable

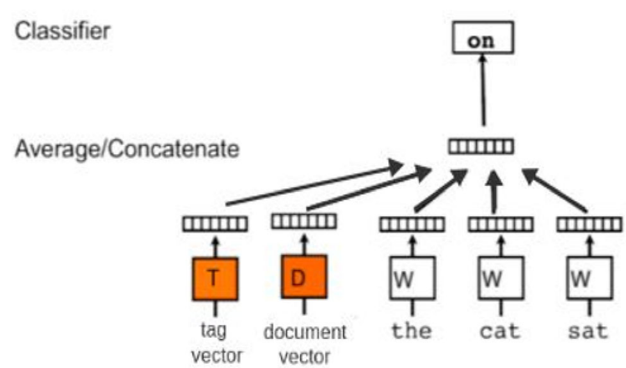

Fig. 2 Doc2Vec model (Ozbay and Alatas 2020)

Table 2 Accuracy of prediction

True Positive (TP)	An unreliable news instance classified as fake
True negative (TN)	An unreliable news instance, classified as reliable
False positive (FP)	A reliable news instance classified as fake
False negative (FN)	A reliable news instance classified as unreliable

Fig. 3 Experiment for classification

```
Input: Data-set, Classifiers
Results: AvgAccuracy, AvgPrecision, AvgF-Score, and AvgAUC
Data set: Kaggle Dataset for Fake News
Classifies ⟵ {DT, RF, GNB, SVM, AdaBoost-DT, AdaBoost-RF, AdaBoost-GNB,
              AdaBoost-SVM, Bagging-DT, Bagging-RF, Bagging-GNB, Bagging-SVM}
All Accuracy Scores ⟵ {}.
All Recall Scores ⟵ {}.
All Precision Scores ⟵ {}.
All F– Scores ⟵ {}.
All AUC Scores ⟵ {}.
for DS ∈ Datasets do
    for Xtrain, Xtest ∈ KFold(nsplits = 10, shuffle = True).split(DS) do
        (Xtrain, Xtest) ← PerformStandardScaler(Xtrain, Xtest);
        ResampledXtrain, ResampledYtrain ← SMOTE(Xtrain, Ytrain);
        for clf ∈ Classifiers do
            clf ← TrainClassifier(clf, ResampledXtrain, XtrainLabels);
            predictions ← predict(cls, Xtest);
            Accurecy ← ComputeAccurecy(predictions, XtestLabels);
            Recall ← ComputeRecall(predictions, XtestLabels);
            Precision ← ComputePrecision(predictions, XtestLabels);
            F-score ← ComputeFmeasure(predictions, XtestLabels);
            AUC ← ComputeAUC(predictions, XtestLabels);
            AllAccuracyScores ← AllAccuracyScores ∪ Accurecy;
            AllRecallScores ← AllRecallScores ∪ Recall;
            AllPrecisionScores ← AllPrecisionScores ∪ Precision;
            AllFScores ← AllFscoreScores ∪ F-score;
            AllAUCScores ← AllAUCScores ∪ AUC;
        end
    end
    AvgAccurecy ← ComputeAvgAccurecy(AllAccuracyScores);
    AvgRecall ← ComputeAvgRecall(AllRecallScores);
    AvgPrecision ← ComputeAvgPrecision(AllPrecisionScores);
    AvgF-score ← ComputeAvgFmeasure(AllFScores);
    AvgAUC ← ComputeAvgAUC(AllAUCScores);
end
return AvgAccurecy, AvgRecall, AvgPrecision, AvgF-score, AvgAUC
```

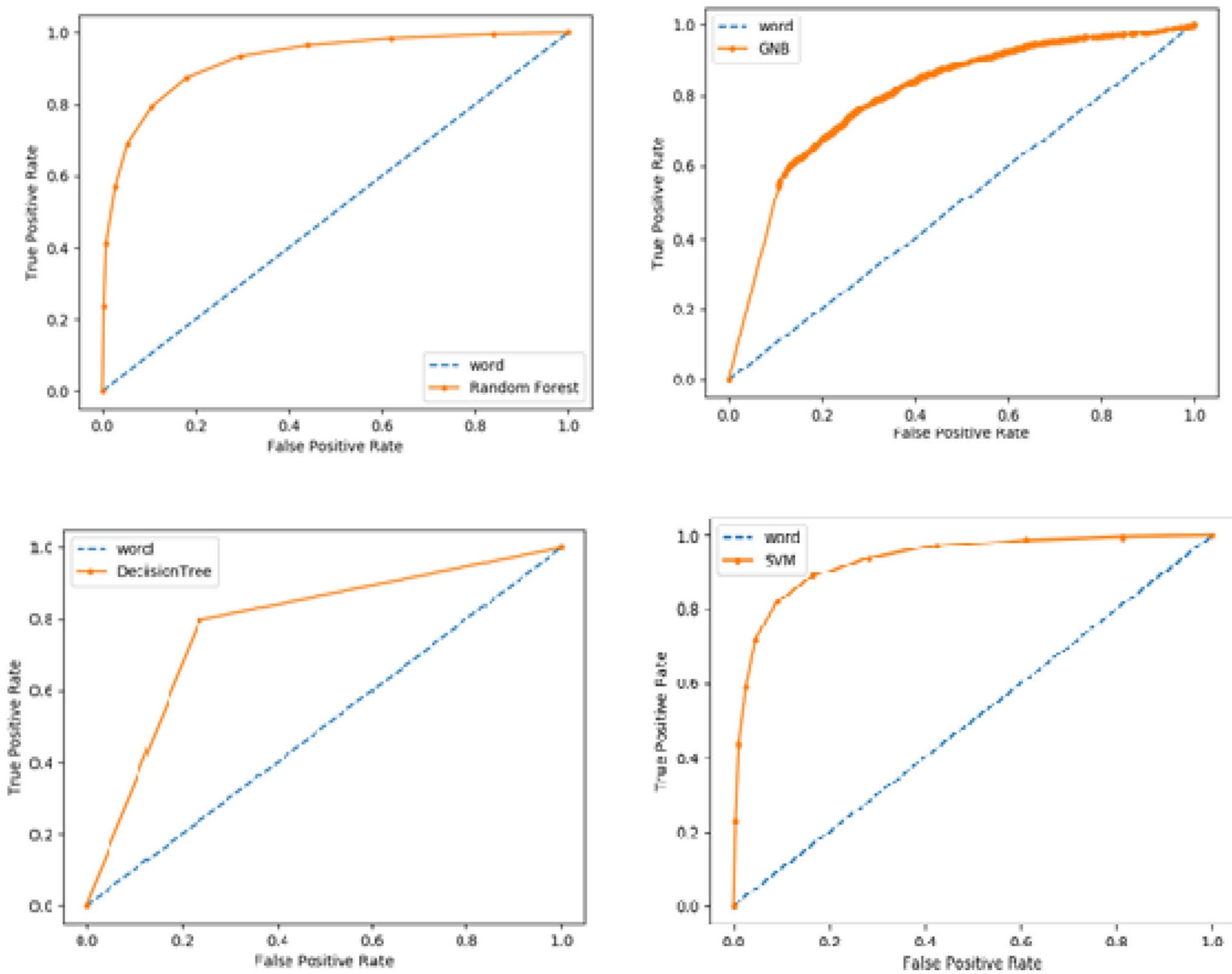

Fig. 4 AUC of Base-Learners (GNB, RF, DT, LinearSVM)

- 0: reliable

3.2 Performance Metrics for Unreliable News

According to Han and colleagues, "Confusion Matrix is commonly used in assessing the classification method" (Han et al. 2012) It is shown in Table 1, TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative. The bigger the TP and TN, the more precise (accurate) the classifier is. All measurements are taken from the Confusion Matrix (Reis et al. 2019). In terms of fake news prediction, the TP, TN, FP, and FN are defined as:

As per rule, the accuracy (additionally called the acknowledgment rate) of a classifier is defined as follows (Google Developers 2019):

$$Accuracy = \frac{TN + TP}{TP + FP + TN + FN} \tag{1}$$

True Positive Rate is also called sensitivity and is defined as follows (Reis et al. 2019):

$$True\ Positive\ Rate = \frac{TP}{TP + FN} \tag{2}$$

False Positive Rate is the proportion of clean software that is wrongly classified as faulty. The FPR can be defined as follows (Han et al. 2012):

$$False\ Positive\ Rate = \frac{FP}{FP + TN} \tag{3}$$

F-score is a harmonic mean of precision and recall and can be defined as follows (Reis et al. 2019):

$$Precision = TP/(TP + FP) \tag{4}$$

$$Recall = TP/(TP + FN) \tag{5}$$

$$F\text{-}Score = (2 * Precision * Recall)/(Precision + Recall) \tag{6}$$

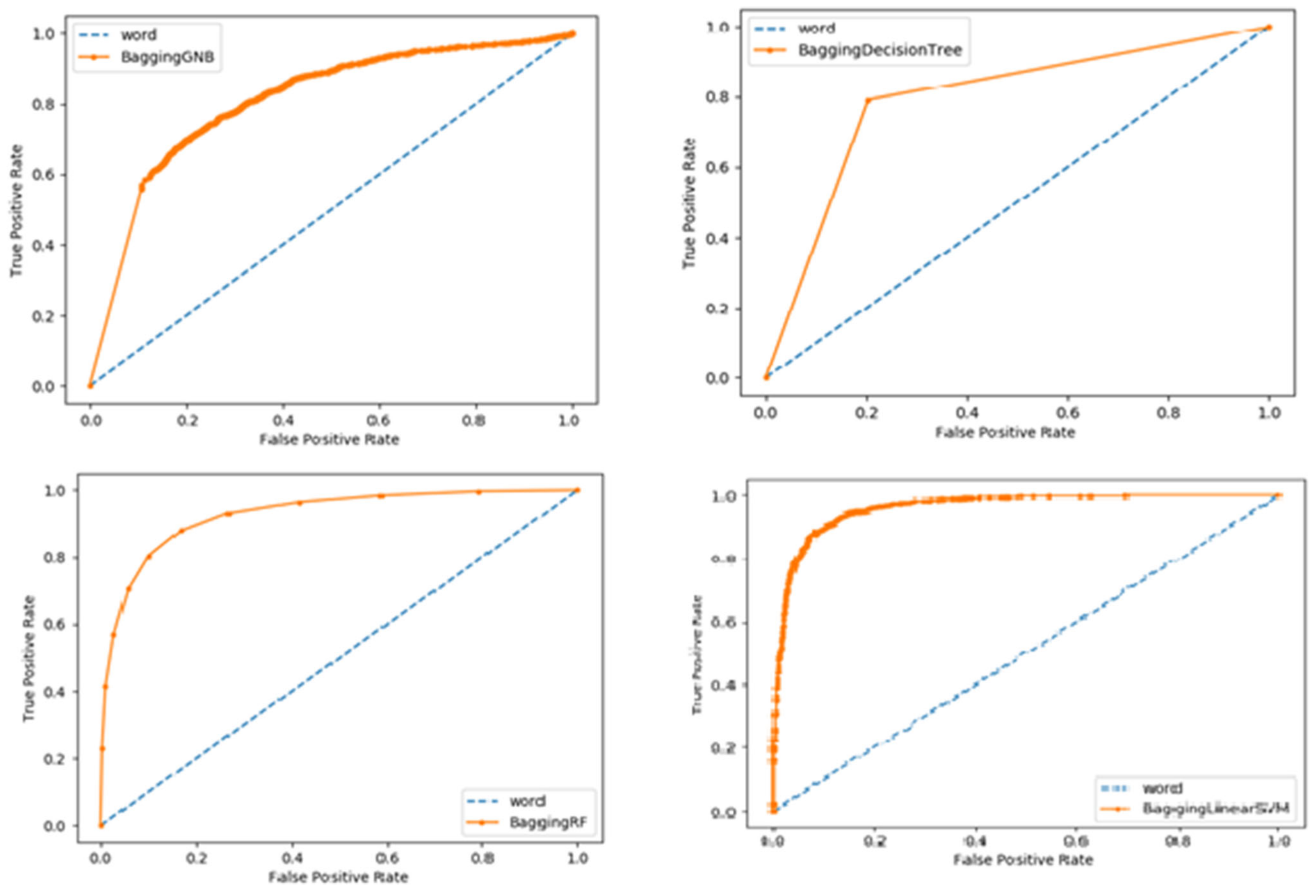

Fig. 5 AUC of Bagging (GNB, RF, DT, Linear-SVM)

AUC is the area under a ROC curve and can be calculated as follows (Reis et al. 2019):

$$AUC = \frac{1 + \mathrm{TPR} + \mathrm{FPR}}{2} \tag{7}$$

3.3 Model design

Embedding (Google Developers 2019) is used for our modelling and is built with the Doc2Vec system. An embedding is a comparatively low-dimensional astronomical into which you can translate high-dimensional vectors. Embedding allows application of machine learning to big inputs like sparse vectors representing words. We perform some essential data pre-preparation before applying Doc2Vec, including eliminating stop words, erasing unique characters and punctuation, and changing the overall content to the lowercase. This provides a comma-separated word list to create a 300-length embedding vector for each article that can be input to the Doc2Vec program. Figure 2 shows that Doc2Vec is a model proposed in 2014, which generates vector representation for words (Quoc and Mikolov 2014). In this structure, each word is mapped to a special vector, represented by a column in a matrix W. The column is recorded by the position of the word in the jargon (vocabulary). Concatenation of all vectors is then utilised as features for the prediction of the next word in a sentence. In Fig. 1, the workflow of the proposed system is given.

4 Experiment

For the experiment, the Kaggle dataset was adopted (Kaggle 2019) as it contains 20,008 training records. For classification, Gaussian Naïve Bayes, Random Forest, Decision Tree, and Support Vector Machine as base learners were used for AdaBoost and Bagging, respectively. The tests were performed using Pycharm in a Python environment. The quality of the classifiers in this study was measured using accuracy, precision, recall, ROC, and F-score of classification. It is important to emphasise that weighted average was used to calculate these metrics. The aim behind the choice of the weighted

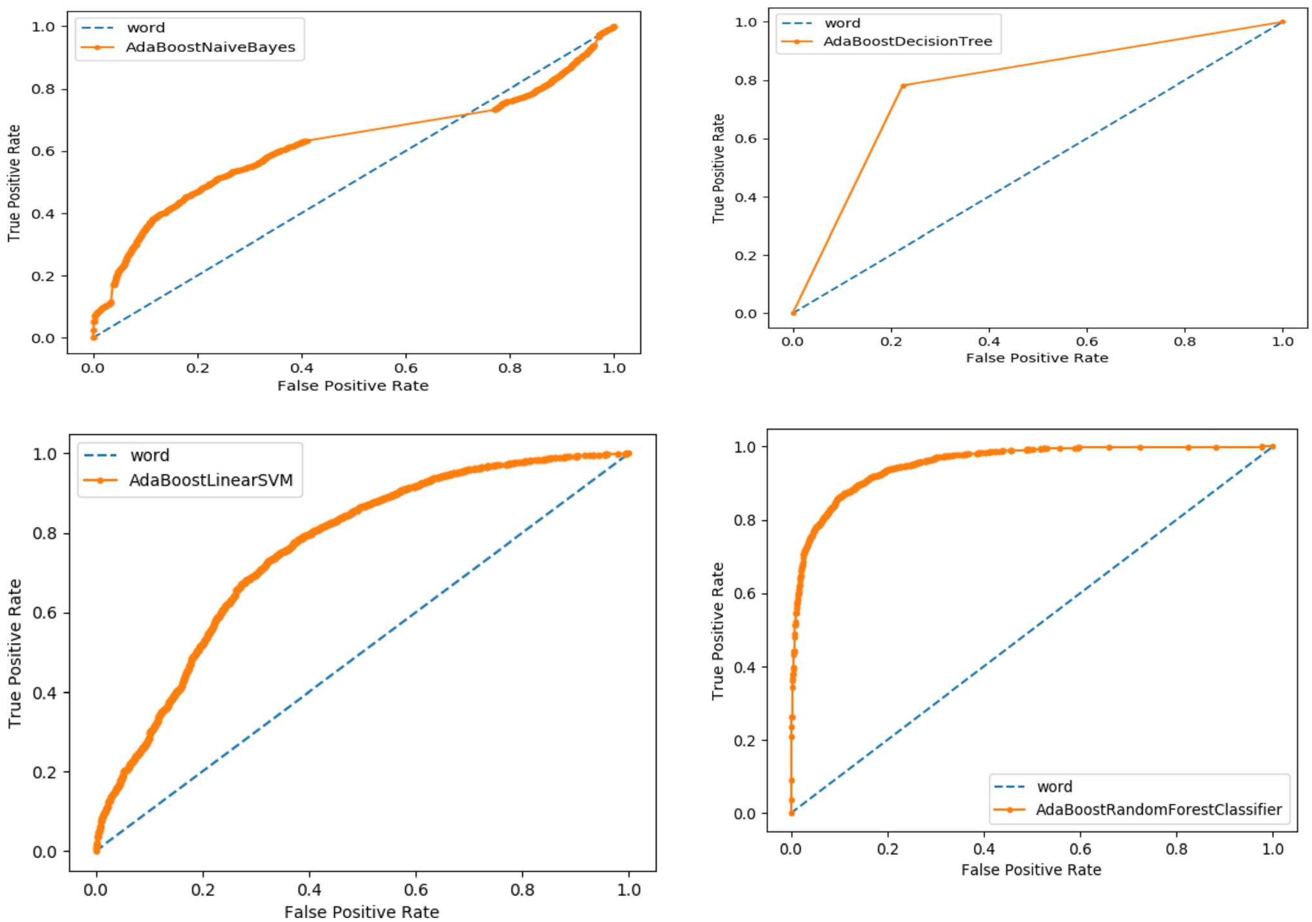

Fig. 6 AUC of AdaBoost (GNB, RF, DT, Linear-SVM)

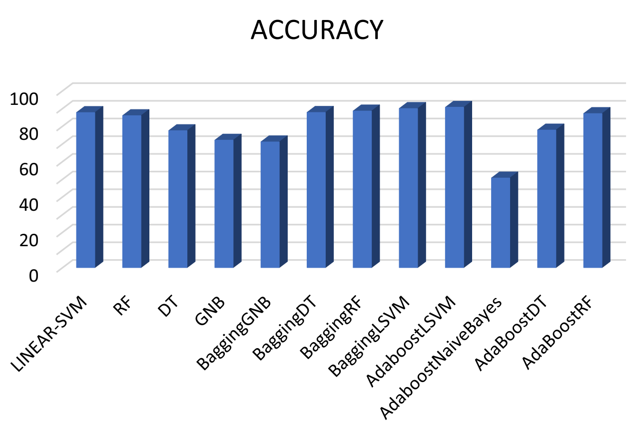

Fig. 7 Accuracy of classifiers

average was to quantify metrics for each group label and to explain the label disparity. Classifier performance was measured with the training dataset based on 10-fold cross-validation. The following algorithm was used for conducting experiments. First, a Kaggle dataset and a list of classifiers were given and then repeated over datasets, as shown in Line 7. As shown in Line 8, the program maintains training and test sets based on 10-fold cross-validation with data shuffling before splitting. For each fold, the loop in Lines 9–20 focused on training the classifiers, obtaining predictions and calculating assessment metrics. The average score was calculated. The process described in Lines 7–28 was iterated through all over the dataset.

5 Results analysis

Experimental results are shown in Table 2 and Figs. 3, 4, 5, 6, 7 and 8. Table 3 shows various classification metrics. The best performance in terms of identifying Fake/Unreliable news was achieved by the AdaBoost-LinearSVM and Bagging-LinearSVM classifiers with 90.7% and 90.02% accuracy, respectively. In general, an AUC of 0.5 suggest no discrimination (mean may be or may not fake information), 0.7–0.8 is considered excellent, and more than 0.9

Fig. 8 Classification results

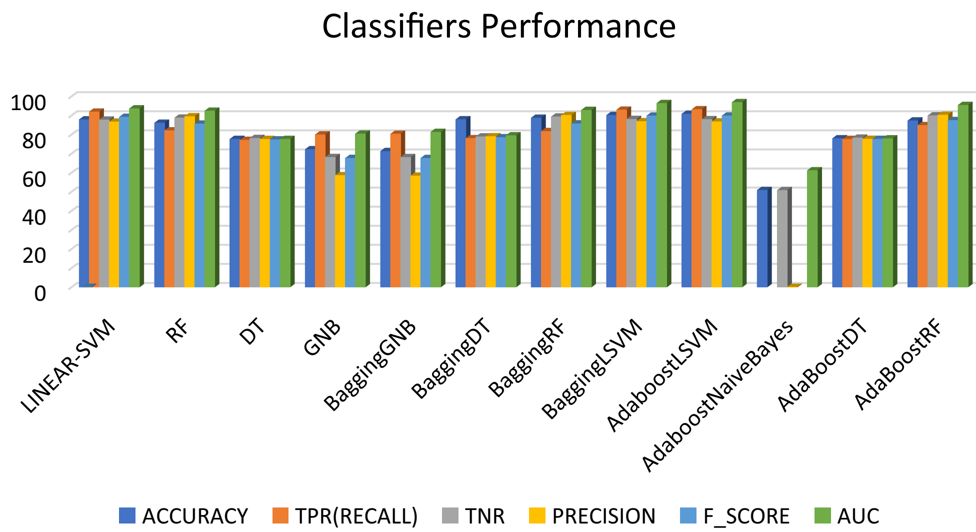

Table 3 Results of various classifiers

Classifiers	Accuracy	TPR (recall)	TNR	Precision	F_SCORE	AUC
LINEAR-SVM	87.74	91.81	87.63	86.56	89.11	94
RF	86	82.01	88.67	89.39	85.52	92
DT	77.51	77.02	78	77.47	77.24	78
GNB	72.12	79.85	68.01	58.5	67.53	80
Bagging-GNB	71.15	80.22	67.98	58.26	67.5	81
Bagging-DT	87.81	77.9	78.8	78.91	78.4	79
Bagging-RF	88.62	81.61	89.24	90.03	85.61	93
Bagging-LSVM	90.02	92.84	87.95	86.8	89.72	96
AdaBoost-LSVM	90.7	93.07	87.86	86.66	89.75	97
AdaBoost-Naïve Bayes	50.73	0	50.68	0	0	61
AdaBoost-DT	77.85	77.49	78.19	77.57	77.53	78
AdaBoost-RF	87.17	84.75	89.83	90.18	87.38	95

is outstanding. The value of AUC shown in Figs. 3, 4 and 5 also supports AdaBoost-LinearSVM and Bagging-LinearSVM with 96% and 97%, respectively, while the Best performance of the classifier is Gaussian Naïve Bayes even using bagging and AdaBoost. The classifiers LinearSVM Bagging-DT, Bagging-RF, and AdaBoost-RF also performed well, with the accuracies of 87.74%, 87.81%, 88.62%, and 87.17%.

6 Conclusion

In this study, analysis of various classifiers was performed, which shows how to detect fake news on social media and to find the sources of fake information. Here, authors calculated the various classification metrics, such as Accuracy, Recall, Precision, F-measure and AUC, revealing that the best performance was achieved by Bagging-LinearSVM and AdaBoost-LinearSVM with accuracies of 90.02% and 90.70%, respectively. The best performers were Bagging-NB and AdaBoost-NB. If to see the value of AUC again bagging and AdaBoost with base learner Linear-SVM have shown the outstanding performance with values for Linear-SVM 94%, Bagging LSVM 97%, Adaboost-RF with 95%, and excellent performance by Decision Tree with value 92%, Bagging RF 93%. In our future work, we plan to apply Deep Learning to classify the fake news; also, to investigate the results on various available datasets like twitter dataset and see how does the accuracy compares to the whole data and form category to another.

References

Aldwairi M, Alwahedi A (2018) Detecting fake news in social media networks. In: The 9th international conference on emerging ubiquitous systems and pervasive networks (EUSPN 2018), vol 141, pp 215–222

Conroy NJ, Rubin VL, Chen Y (2015) 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, American Society for Information Science, Silver Springs, MD, USA, pp 82:1–4. http://dl.acm.org/citation.cfm?id=2857070.2857152

Dungs S, Aker A, Fuhr N, Bontcheva K (2018) Can rumour stance alone predict veracity? In: Proceedings of COLING 2018, the 27th international conference on computational linguistics, pp 3360–3370

Enayet O, El-Beltagy SR (2017) Niletmrg at semeval-2017 task 8: determining rumor and veracity support for rumours on twitter. In: Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017), pp 470–474

Ferreira W, Vlachos A (2016) Emergent: a novel data-set for stance classification. In: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, pp 1163–1168

Gardiner B (2015) You'll be outraged at how easy it was to get you to click on this headline. Wired. https://www.wired.com/2015/12/psychology-of-clickbait/. Accessed June 2020

Google Developers. https://developers.google.com/machine-learning/crash-course/embeddings/. Accessed Nov 2019

Han J, Kamber M, Pei J (2012) Data mining concepts and techniques, 3rd edn. Morgan Kaufmann, Burlington

Hardalov M, Koychev I, Nakov P (2016) In search of credible news. In: International conference on Artificial intelligence: methodology, systems, and applications, pp. 172–180. Springer

Hassid J (2011) Four models of the fourth estate: a typology of contemporary Chinese journalists. China Q 208:813–832. https://doi.org/10.1017/S0305741011001019

Jin Z, Cao J, Zhang Y, Luo J (2016) News verification by exploiting conflicting social viewpoints in microblogs. In: AAAI, pp. 2972–2978

Kaggle. https://www.kaggle.com/c/fake-news/data. Accessed 3 Nov 2019)

Marchi R (2012) With Facebook, blogs, and fake news, teens reject journalistic objectivity. J Commun Inq 36:246–262. https://doi.org/10.1177/0196859912458700

Markowitz DM, Hancock JT (2014) Linguistic traces of a scientific fraud: the case of Diederik Stapel. PLoS ONE 9(8):e105937

Ozbay F, Alatas B (2020) Fake news detection within online social media using supervised artificial intelligence algorithms. Phys A Stat Mech Appl 540:123174

Popat K, Mukherjee S, Strötgen J, Weikum G (2017) Where the truth lies: explaining the credibility of emerging claims on the web and social media. In: Proceedings of the 26th international conference on world wide web companion. International World Wide Web Conferences Steering Committee, pp 1003–1012

Quoc L, Mikolov T (2014) Distributed representations of sentences and documents. https://arxiv.org/abs/1405.4053

Reis JCS, Correia A, Murai F, Veloso A, Benevenuto F (2019) Explainable machine learning for fake news detection. In: Proceedings of the 10th ACM conference on web science, pp 17–26

Reis JCS, Correia A, Murai F, Veloso A, Benevenuto F (2019b) Supervised learning for fake news detection. IEEE Intell Syst 34(2):76–81

Rubin VL, Chen Y, Conroy NJ (2015) 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, vol 83. American Society for Information Science, Silver Springs, MD, USA. pp 1–4. http://dl.acm.org/citation.cfm?id=2857070.2857153

Smith J, Leavitt A, Jackson G (2018) Designing new ways to give context to news stories. https://medium.com/facebook-design/designing-new-ways-to-give-context-to-news-stories-f6c13604f450

Spicer RN (2018) Lies, damn lies, alternative facts, fake news, propaganda, pinocchios, pants on fire, disinformation, misinformation, post-truth, data, and statistics. Springer, Cham, pp 1–31. https://doi.org/10.1007/978-3-319-69820-5_1

Zuckerberg M (2016) Facebook post. https://www.facebook.com/zuck/posts/10103253901916271