



# User Response-Based Fake News Detection on Social Media

Hailay Kidu, Haile Misgna, Tong Li<sup>(✉)</sup>, and Zhen Yang

Faculty of Information Technology, Beijing University of Technology, Beijing, China  
{hailay.kidu,haile.misgna}@mu.edu.et, {litong,yangzhen}@bjut.edu.cn

**Abstract.** Social media has been a major information sharing and communication platform for individuals and organizations on a mass scale. Its ability to engage users to react to information posted on this media in the form of like, share, and comment made it a preferable information sharing platform by many. But the contents posted on social media are not filtered, fact checked or judged by an editorial body like any traditional news platform. Therefore, individuals, institutions and communities who consume news from social media are vulnerable to misinformation by malicious authors. In this work, we are proposing an approach that detects fake news by investigating the reaction of users to a post composed by malicious authors. Using features extracted by bag-of-words model and TF-IDF from text based replies (comments), and visual emotion responses in the form of categorical data, we built models that predicted news as fake or real. We have designed and conducted a series of experiments to evaluate the performance of our approach. The results show the proposed approach outperforms the baseline in all the six models. In particular, our models from random forest, logistic regression, and XGBoost algorithms produce a precision of 0.97, a recall of 0.99 and an F1 of 0.98.

**Keywords:** Fake news detection · Users' responses · Machine learning · Deep learning

## 1 Introduction

The capability of social media technologies to communicate and disseminate information on a massive scale, high speed, low cost, easy access and freedom to publish anything contributed in drawing mainstream news consumers and publishers into social media, such as Twitter, Facebook, blogs and others. The freedom to publish anything on those platforms is a double edged sword. The contents posted on those platforms are not filtered, fact checked or judged by an editorial body. Most people do not verify the source of the news before they read and share it, which can lead to the propagation of fake news quickly and may lead to it going viral. Individuals, institutions and communities who consume news from those platforms are vulnerable to malicious authors whose intention is misinforming readers.

The expression “fake news” is popularized during the 2016 US elections [20]. Since then researchers in journalism and artificial intelligence are working on defining fake news and how to identify it automatically using the help of computer algorithms as early as possible. Allcott and Gentzkow [6] define fake news as intentionally written to mislead news consumers with information which are verifiably false.

The seriousness of the threats and fears caused because of fake news has a high impact in the integrity of journalism and politics. To mitigate those threats and fears Nakamura et al. [18] and Wang [30] proposed a benchmark datasets for fact checking. Abbasi and Liu [2], and Sitaula et al. [27] claim assessing the association of users with fake news is an important feature in detecting fake news. Choudhary and Arora [9] proposed a linguistic feature approach that employs syntax, sentiment, grammatical and readability evidence to characterize fake news from textual contents to design a set of features. Shu et al. [25] explores auxiliary information from tri-relationship, a relationship between publishers, users, and news content, to improve the process of fake news detection.

In all those works, features extracted from news content are used to detect fake news. Even though users’ responses are honest reactions to a news content with less intention to deceive others, we found few works that considered users’ reactions in the task of fake news detection. The works which considered user responses treated features extracted from users’ reactions as auxiliary source of information. On social media platforms, like Twitter and Facebook, there is a high degree of user engagement on posted news. Comments are written and emotions are expressed against or in support of the post.

In this research paper, we investigated fake news detection on Twitter from the users’ responses presented in the form of text and visual emotional reaction. To our knowledge there is no comparative study done on fake news detection approaches; news content-based approach and social context approach. In this work, we are proposing the use of users’ responses which is a type of social context as a main source of information to detect fake news rather than using them as an auxiliary source of information. We proposed a new way of looking at the users’ responses information in the study of fake news detection, and we did comparative analysis of fake news detection on users’ responses and news content using machine learning algorithms.

The remaining sections of this article are structured as follows. In Sect. 2 related works are covered. Section 3 discusses the methodologies. Experimental results and analysis are presented in Sect. 4. In Sect. 5 discussion on the overall results is discussed. The last Sect. 6 makes concluding remarks and discusses future works.

## 2 Related Work

Fake news identification using computer algorithms has been not trivial. Shu and Liu [22] presented four unique challenges of fake news on social media: (1) it is not simple to detect fake news simply based on content, (2) the volume,

variety and veracity characteristics of the social media data, (3) the background of social media users, and (4) the easiness to create malicious accounts. These challenges motivate researchers to study the mitigation of fake news propagation by detecting them as early as possible in order to prevent tremendous negative political and social impacts.

Shu and Liu [22] classify existing fake news detection approaches into two broad categories: fake news detection from news contents and fake news detection from social contexts.

## 2.1 News Content-Based Approach

News contents contain a great deal of information that can tell whether a given news is fake or real. The features that characterize the news content can be the source of the news, the headline, the main text, and visual information embedded in the news. From those features more discriminative characteristics of fake news are built. This approach is the most studied approach. A number of works employ techniques from natural language processing and machine learning, and apply it on the news content [3, 7, 11, 12, 28, 29].

Wang [30] and, Nakamura et al. [18] proposed benchmark datasets for fake news detection. Nakamura et al. [18] present a large-scale multimodal fake news dataset which contains over 1 million samples containing text, image, metadata, and comments from a highly diverse set of resources. The dataset is a multiple labeled dataset with 2-way, 3-way, and 6-way classification. In addition, the dataset utilizes image data as evidence for text truthfulness or text data for image truthfulness.

Abbasi and Liu [2], and Sitaula et al. [27] claimed assessing credibility of users has a significant role in detecting fake news. Abbasi and Liu [2] proposed CredRank algorithm to measure users' credibility on social media. The credibility score is built from the behaviour of users on social media that is posted as a news content. The authors argue that a user with high credibility score is less likely to propagate fake news detection. Studying the role of user profile helps in detecting who will likely share a fake news [26].

Choudhary and Arora [9] proposed a linguistic model to find out the properties of news contents that will generate language-driven features. This model extracts syntactic, grammatical, sentimental, and readability features of particular news. The linguistic feature-driven model achieved the average accuracy of 86% for fake news detection and classification. Granik and Mesyura [11] proposed a simple approach for fake news detection using naive Bayes classifier on news contents, and achieved a classification accuracy of approximately 74%.

Khan et al. [15] argues that most fake news detection algorithms are trained in politics dataset and this will result in producing biased models. The authors combined three different datasets with diverse topics and investigated the performance of different machine learning models. Ahmed et al. [4] proposed n gram models for fake news detection and they applied TF-IDF for feature extraction. They conducted a comparative analysis on six machine learning models and they got 92% accuracy with linear SVM classifiers. Aslam et al. [7] proposed

an ensemble-based deep learning model to classify the news into fake or real on LIAR dataset. NLP techniques were applied to extract text features from news content. A deep learning model, Bi-LSTM-GRU-dense model, on news content attributes achieved a better performance result.

Horne and Adali [14] discussed systematic, stylistic and other content differences between fake and real news. In conducting their fake news study, they investigated three separate data sets. They also include satire as a type of fake news that relies on absurdity other than sound arguments to make claims, but explicitly identifies itself as satire. Shu et al. [23] proposed a general data mining framework for fake news detection which includes two phases: (i) feature extraction and (ii) model construction. This work presented the narrow and broad definitions of fake news and clear direction from characterization to detection.

Yang et al. [31] proposed an unsupervised learning framework, which utilizes a probabilistic graphical model to model the truths of news and the users' credibility. An efficient collapsed Gibbs sampling approach is proposed to solve the inference problem. They conducted experiments on two real-world social media datasets, LIAR and BuzzFeed. Their experimental results demonstrated the effectiveness of the proposed framework for fake news detection on social media.

## 2.2 Social Context Approach

In this approach, the engagement of users to a news content via social media generates supplementary information that likely enhances the content-based models. A number of research articles address the challenge of fake news detection from different perspectives.

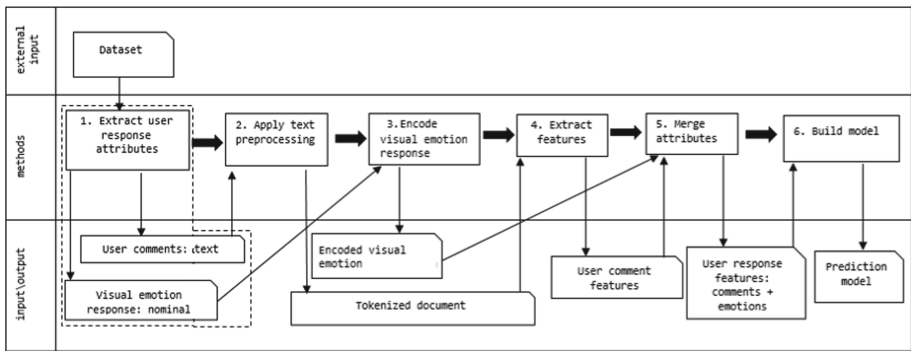
Shu et al. [24] argued that the performance of models built on news content only is not satisfactory, and it is suggested to incorporate user social engagements as auxiliary information to improve the fake news detection task. They constructed real-world datasets measuring users' trust level on fake news and select representative groups of both "experienced" and "naïve" users. And they performed a comparative analysis over explicit and implicit profile features between those user groups.

News content-based approach is the well researched approach. But targeting news contents for detection of fake news raises too many challenges. One of the main challenges for this approach is that news contents are carefully crafted to deceive readers. So, we argue that it is not trivial to target the mitigation of fake news detection from a news content angle. Shu et al. [25] discussed the helpfulness of social context approach in providing auxiliary information that may result in having better performance results. But we presented a different approach of using social context information in the process of fake news detection, i.e., using the users' responses information as the main data input in detecting fake news. The responses of users' contain rich information and the input data is presented in the form of text and visual emotion. To our knowledge, we have not found an article that addresses fake news detection from the perspective of

users’ responses. In this work, we proposed the user response-based approach for automatically classifying news as fake or real on Twitter.

### 3 Methodologies

The major question this paper addresses is whether we can automatically detect fake news by the textual and visual response of users that outperforms the news content approach. In this work, we proposed a new approach that targets users’ reactions in detecting fake news. The proposed approach involves different tasks to shape the input before being used for model production. These tasks include data preprocessing, feature extraction and model production. Figure 1 and the following sections show the methods employed in this work.



**Fig. 1.** A users’ response-based fake news detection methodology. The first row shows the input data. The second row presents all the set of tasks employed in the proposed approach. And the last row presents the intermediate outputs and inputs of the tasks in row two.

#### 3.1 Extract User Response Attributes

The dataset has attributes like idx, context\_idx, text, reply, categories and mp4. In this stage, attributes that define users’ response are selected. The attributes reply, mp4 and categories are attributes that show users’ engagement to a tweet in the dataset we are experimenting with. Categories and mp4 represent the same type of information; categories attribute is a description of the GIF files in the mp4 attribute. Therefore, we considered categories attribute which represents the visual information and reply attributes that represents users comment as user response attributes for this work.

### 3.2 Apply Text Preprocessing

Before extracting features from the textual data, the data needs to be cleaned to a certain level to increase the performance of the models. Therefore, we applied data cleaning and transformation operations on the textual data, such as numbers and special characters removal, stopwords removal, tokenization, and stemming and lemmatization.

- **Removal of special and single characters:** symbols like ?, <, >, +, -, /, \*, !, @, #, \$, %, ;, &, (, ), {, }, [, ], ‘, “, ;, :, ~, and more are removed. In addition to that, numbers and single characters are also removed because it is assumed that they don’t have much significance in the fake news identification process.
- **Stopwords removal:** words that are frequently found in documents that have less discriminating power are removed to increase the performance of the classifiers. We used the English stopwords list from NLTK [16] to conduct this operation.
- **Tokenization:** by normalizing all the words in the document into lower case, tokenization is applied to the text features. The tokenization operation transforms the text entry into a list of words that can be used as a feature in the model building task.
- **Word stemming and lemmatization:** WordNetLemmatizer from NLTK is used to transform the text into its lemma form. This operation will reduce the dimension of the features by transforming inflected and derivational words.

### 3.3 Encode Emotions

The visual emotional information is described as categories attribute. This attribute is a nominal data presented in the form textual data format. Therefore, we encode this categorical data using one-hot encoding to transform the nominal textual input into a binary incidence data input. The one-hot encoding determines the presence or absence of emotions at every entry of the user. The encoded data shows an emotional reaction of user to a post. This encoding creates a categories matrix,  $C$ , that represents emotion  $e$  expressed by user  $u$ .

$$C_{u,e} = \begin{cases} 1, & \text{if the emotion } e \text{ is expressed by user } u \\ 0, & \text{otherwise} \end{cases}$$

The entries are encoded 1 if the emotion is expressed by a given user, otherwise the entries are 0. In this matrix the rows represent the users and the columns represent the emotions.

### 3.4 Extract Features

The model building process and performance is highly dependent on the set of features used for the identification of fake news. This work mainly focuses on two

attributes of the dataset, namely reply and categories. The categories attribute is nominal data whereas the reply is textual data. For this work, we used state-of-the-art feature extraction models for text classification, n-gram models. This model is used for text classification in short texts [10, 11] and large contents [4, 5].

**BoW Model with TF-IDF Weighting Scheme for Feature Extraction:** Bag-of-words (BoW) model is used to extract features from the reply attribute which is a text entry. The process of feature extraction starts from tokenization of the reply entries into a word vector. And then, we applied a TF-IDF weighting scheme to measure the relevance of each term.

$$tf(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \quad (1)$$

The above formula is read as, term frequency of  $t$  in document  $d$ ,  $tf(t,d)$ , is the frequency of term  $t$  in document  $d$ ,  $f_{t,d}$ , divided by the sum of the frequencies of all the words in the document,  $\sum_{t' \in d} f_{t',d}$ .

$$idf(t, D) = \log \frac{N}{|\{d \in D : t \in d\}|} \quad (2)$$

Inverse document frequency of a term  $t$  in document  $d$  is the log of total number of collections in the corpus  $N$  divided by the number of documents  $d$  that contain the term  $t$ . So then, TF-IDF is the product of the two equations.

$$TF - IDF = tf(t, d) * idf(t, D) \quad (3)$$

### 3.5 Merge Attributes

The user's reaction is defined by features extracted from the comments and visual emotions. We extracted 100 features from the comments and 44 features from the visual emotions. Both data are presented in matrix form. The first matrix represents features extracted from comments. The entries of this matrix are TF-IDF of each feature expressed by a user. The second matrix represents the emotion of users expressed towards a news post. The entries of this matrix represent the incidence of emotions. Each entry shows whether a given user expressed an emotion to a news post. By merging both matrices using NumPy [13] `hstack` function, we created a matrix that has 100 features from the text entry and 44 features from the nominal data. Therefore, a new form of data with 144 features is generated and used as an input to the algorithms to create fake news prediction models.

### 3.6 Build Model

To test our approach, we build models from six machine learning algorithms, namely random forest, logistic regression, support vector machine (SVM) and

multinomial naive Bayes, XGBoost [8] and RNN, and investigate how the proposed approach performed. We used scikit learn API for building models from random forest, logistic regression, support vector machine, and multinomial naive Bayes [19]. For RNN, we used tensorflow to build as well as to test our model [1]. We trained our model in two types of input data; text and nominal data. For the RNN model, we used word embedding [17] as input and features from categories attribute where the output of the word embedding is concatenated before going to the activation layer. For the other algorithms, TF-IDF vectors of the reply attribute and feature vectors from categories attribute are concatenated and given as an input.

## 4 Experimental Results and Analysis

We conducted experiments on the proposed approach for fake news detection and compared it with a baseline, a method that detects fake news from news contents using a bag-of-words model. We did comparisons on 6 machine learning algorithms and the proposed approach outperformed in all the algorithms. In this work, we will answer the following research questions (RQ):

- RQ 1: Will users’ responses provide more helpful clues than news content in fake news classification?
- RQ 2: Which one of the machine learning models performs best in detecting fake news from users’ responses?

### 4.1 Dataset

This work uses a publicly available dataset prepared for the challenge of Fake-EmoReact 2021 Challenge<sup>1</sup>. The dataset is collected from Twitter and the records with at least one GIF response are considered in the preparation of the dataset. The challenge organizers prepared the dataset in 3 different JSON files;train.json, eval.json and dev.json. The train.json file is the only labeled dataset but both eval.json and dev.json are not labeled, they are the holdout data. train.json contains 168,521 entries and 6 features. The 31,799 (19%) entries are labeled as real and 136,722 (81%) of them are labeled as fake. There are only 227 unique tweets. The dataset has idx, context\_idx, mp4, text, reply and categories attributes.

- **idx** is an identifier of a tweet
- **text** is the original tweet
- **context\_idx** is an identifier of a reply for a given tweet
- **reply** is user response in the form of text
- **mp4** is a GIF user response that represents visual emotional reactions
- **categories** is a textual description of mp4 attribute

<sup>1</sup> <https://sites.google.com/view/covidfake-emoreact-2021/>.



## 4.2 Experimental Settings

To show detecting fake news from users' reactions is better than news content post, we conducted a comparison of fake news detection from news content (user tweet) and user emotion reaction in the following experimental settings and evaluation protocols.

**Parameter Settings:** The parameters are chosen after conducting many iterations to fine-tune the trade-off between speed and performance. The settings used in conducting the experiments are stated as follows:

- TF-IDF: One hundred unigram features are generated from the reply (text) attributes. The minimum document size is five and the maximum document size is 70%.
- Word embeddings: In this feature engineering task, we consider only 100 features. In other words, the vocabulary size of the embedding layer is 100.
- Random forest: The number of trees used to estimate the label is 100.
- Logistic regression, linear support vector machine, and naive Bayes we used the default parameters of scikit learn.
- RNN: The number of batches for training are 128, the number of iteration (epochs) are 25 and ADAM optimizer is used.

**Evaluation Protocol:** To perform the experiments, we first trained six algorithms in the training.json data. We did not split the dataset into training and testing data. We used train.json (training data from the challenge organizers) for training and dev.json (validation data from the challenge organizers) for testing. Every experimental results included in this work are results collected after they are tested on the dev.json file. This file is deliberately prepared by the organizers to test the performance of the models on held out data. All our performance scores are collected from CodaLab<sup>2</sup> after submitting the results of our models. And the performance of the models is presented using precision, recall and f-measure.

**Baseline:** We compare our proposed approach against approaches that detect fake news from news content using linguistic features.

- Detecting fake news from news content (**News content**): Ahmed et al. [4] proposed a methodology that detects fake news from news content using BoW model. We used the same feature extraction methodology for the classical machine learning algorithms.

---

<sup>2</sup> <https://competitions.codalab.org/competitions/30741>.

**Table 1.** The comparison of the proposed approach (users’ responses) against the baseline (news content) using logistic regression, random forest, NB, linear SVC, XGBoost, and RNN.

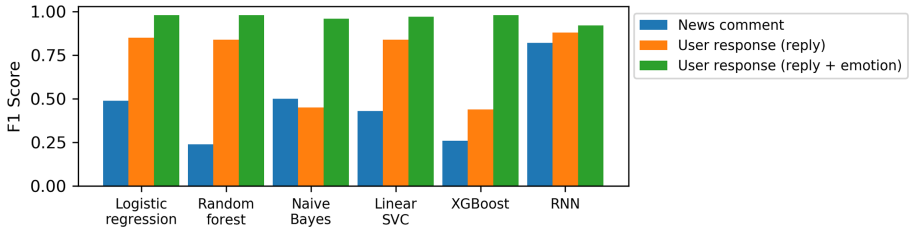
Algorithms	Metrics	News content [4]	User response (proposed)
Logistic regression	Precision	0.57	<b>0.97</b>
	Recall	0.6	<b>0.99</b>
	F1-score	0.49	<b>0.98</b>
Random forest	Precision	0.56	<b>0.97</b>
	Recall	0.52	<b>0.99</b>
	F1-score	0.24	<b>0.98</b>
NB	Precision	0.61	<b>0.97</b>
	Recall	0.52	<b>0.95</b>
	F-score	0.5	<b>0.96</b>
Linear SVC	Precision	0.55	<b>0.97</b>
	Recall	0.56	<b>0.98</b>
	F-score	0.43	<b>0.97</b>
XGBoost	Precision	0.57	<b>0.97</b>
	Recall	0.53	<b>0.99</b>
	F-score	0.26	<b>0.98</b>
RNN	Precision	0.79	<b>0.89</b>
	Recall	0.91	<b>0.96</b>
	F-score	0.82	<b>0.92</b>

### 4.3 Performance Comparison (RQ1)

Table 1 shows the performance report of the six models on news content and users’ responses approaches. This summary report shows the score of the two approaches using precision, recall and f1-score. According to our experiment, the proposed approach shows better performance. The users’ response data is built from two different inputs, reply and GIF responses. Since the reply attribute is a text in nature, we conducted an experiment to see if it can produce a different result when it is compared with the data with the same data type. As it is shown on Fig. 2, the users’ response from reply, and merged data of both reply and emotional information outperformed the baseline. RNN outperforms all the other classical machine learning models on news comment and user response(reply) data. This showed that word embedding is better in extracting discriminative features from text input data than the simple bag-of-words model.

We demonstrated the betterment of detecting fake news from users’ comments (replies) and the whole users’ responses (reply and visual emotional) information. The comparisons between news comment and users’ responses on the textual input showed that more interesting textual features are extracted from users’ responses.

However, on the proposed approach column the other models improve significantly when the categories attribute is merged with the replies attribute. The experimental results show that logistic regression, random forest and XGBoost models performed better with mixed data. The use of GIF emotions helped the

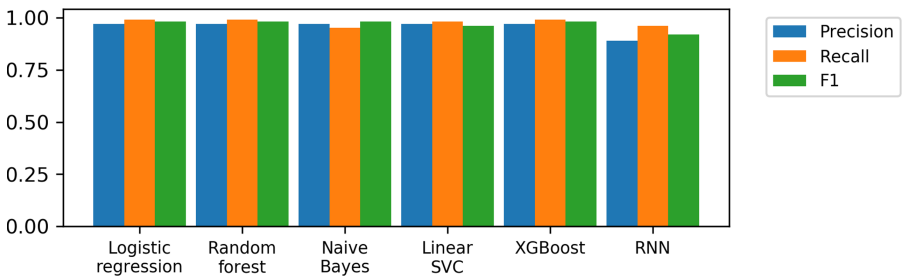


**Fig. 2.** Comparison of users’ response-based vs. news content fake news detection (F1 score).

naive Bayes algorithm to jump from 0.50 F-1 score to 0.96. Similar increment patterns are shown for all the models with mixed data. In general, the use of emotions increased the performance of all the six algorithms. But the increase in performance based on categories is not that much on RNN. We have observed a similar result as [21] where the adding of LIWC features to features extracted from news content using word embeddings does not contribute any significant increase in performance. Therefore, we draw the same conclusion as [21] that the RNN model already learned most of the features from the text data and most of the features from categories attribute are redundant.

#### 4.4 Which Algorithm Fits Best? (RQ2)

As it is seen on Fig. 3, our proposed approach is tested using five classical machine learning algorithms and one deep learning algorithm. Based on the experimental results, it has been shown that the ability of learning more patterns from the visual emotional information is demonstrated by the classical machine learning models. The merging of those features from users’ textual and visual emotion reactions highly boosted the performance of the classical machine learning algorithms whereas the performance of RNN stayed with no significant increase. Logistic regression, random forest and XGBoost algorithms build better models using the basic feature extraction techniques.



**Fig. 3.** An experimental result of users’ response-based fake news detection.

## 5 Discussion

Most of the works on fake news detection that used Twitter dataset mainly target the tweets attribute. The tweets are the news contents that are intentionally composed to misinform followers. Therefore, it is not easy to distinguish fake from real by intensively studying the tweets (news contents). But replies are written from the angle of reaction given to a tweet. Replies contain actual emotion of followers towards the news content. A number of followers express a range of emotions in support and against the original news content.

The contents of replies are rich in content. And the contents are presented in the form of text and visual information. This information has been used as auxiliary information in the process of fake news detection. In this work, we want to highlight how much it is significant to consider users' responses in the study of fake news detection. To shed a light on the significance of this type of information, we compared it against news content where in major works used to detect fake news. We used this as a baseline for this work.

We used simple feature engineering techniques for feature extraction from the text, n-gram model using TF-IDF weighting scheme. By applying this method on both tweet and replies (text), we have demonstrated that the replies (text) outperformed the tweets in the process of fake news detection. Users' responses that contain both textual and emotional replies showed the best performance result in comparison with replies (text) and tweets (news content).

To demonstrate the users' responses contributes more interesting features than tweets in deep learning as well, we tested it using word embeddings and RNN. The deep learning model showed a good result in replies with textual and visual information. But the performance on textual replies is nearly the same with the replies that contain both textual and visual information. Therefore, the model did not learn not much new pattern from the visual information.

In summary, our investigation showed an important approach for the task of fake news detection from the perspective of users' responses. Specifically, we have observed classifiers that depend on visual emotion information perform consistently good on all the algorithms we tested. Figure 2 demonstrated the incorporation of visual information led to a better performance in the task of fake news detection. As it is shown on Fig. 3, random forest, logistic regression and XGBoost models showed a better result than the other models. The visual information presented in the form of categorical data contributed a significant factor of increase in performance on all the classical machine learning algorithms but not in RNN.

## 6 Conclusion and Future Work

Automatic fake news detection employs different techniques and methods from natural language processing and machine learning. Feature generation and training models are the major works in the process of fake news classification. In this paper, we addressed the task of automatic fake news identification from the perspective of who provides more clues. We proposed an approach of addressing

fake news from the perspective of followers where their responses are manifested in the form of textual and visual emotion data. We built different classification models that showed users' responses contain more discriminative information in the task of fake news detection.

In this work, we only investigated the contents of the users' responses. We addressed the problem from a text classification perspective where simple syntactic approaches are used. For future work, we will explore more text classification algorithms. And it is useful to investigate domain-specific features, linguistic features, post-based detection, and network-based detection feature extraction methods. In algorithmic-wise, investigating transfer learning feature generation models, such as BERT, GloVe, etc. will be useful to extend our approach.

**Acknowledgments.** This work is partially supported by the National Natural Science of Foundation of China (No. 61902010, 61671030), the International Research Cooperation Talent Introduction and Cultivation Project of Beijing University of Technology (No. 2021C01), and the Project of Beijing Municipal Education Commission (No. KM202110005025).

## References

1. Abadi, M., et al.: TensorFlow: a system for large-scale machine learning. In: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 2016), pp. 265–283 (2016)
2. Abbasi, M.-A., Liu, H.: Measuring user credibility in social media. In: Greenberg, A.M., Kennedy, W.G., Bos, N.D. (eds.) SBP 2013. LNCS, vol. 7812, pp. 441–448. Springer, Heidelberg (2013). [https://doi.org/10.1007/978-3-642-37210-0\\_48](https://doi.org/10.1007/978-3-642-37210-0_48)
3. Ahmad, I., Yousaf, M., Yousaf, S., Ahmad, M.O.: Fake news detection using machine learning ensemble methods. *Complexity* **2020** (2020). <https://doi.org/10.1155/2020/8885861>
4. Ahmed, H., Traore, I., Saad, S.: Detection of online fake news using N-gram analysis and machine learning techniques. In: Traore, I., Woungang, I., Awad, A. (eds.) ISDDC 2017. LNCS, vol. 10618, pp. 127–138. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-69155-8\\_9](https://doi.org/10.1007/978-3-319-69155-8_9)
5. Ahmed, H., Traore, I., Saad, S.: Detecting opinion spams and fake news using text classification. *Secur. Priv.* **1**, 1–15 (2018)
6. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. *J. Econ. Perspect.* **31**, 211–236 (2017)
7. Aslam, N., Ullah Khan, I., Alotaibi, F.S., Aldaej, L.A., Aldubaikil, A.K.: Fake detect: a deep learning ensemble model for fake news detection. *Complexity* **2021** (2021). <https://doi.org/10.1155/2021/5557784>
8. Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al.: XGBoost: extreme gradient boosting. *R Package Version 0.4-2* **1**(4), 1–4 (2015)
9. Choudhary, A., Arora, A.: Linguistic feature based learning model for fake news detection and classification. *Expert Syst. Appl.* **169**, 114171 (2021)
10. Dilrukshi, I., De Zoysa, K., Caldera, A.: Twitter news classification using SVM. In: 2013 8th International Conference on Computer Science & Education, pp. 287–291. IEEE (2013)

11. Granik, M., Mesyura, V.: Fake news detection using Naive Bayes classifier. In: 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKR-CON), pp. 900–903. IEEE (2017)
12. Hajek, P., Barushka, A., Munk, M.: Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining. *Neural Comput. Appl.* **32**(23), 17259–17274 (2020). <https://doi.org/10.1007/s00521-020-04757-2>
13. Harris, C.R., et al.: Array programming with NumPy. *Nature* **585**, 357–362 (2020)
14. Horne, B., Adali, S.: This just. In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar To Satire Than Real News. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 11 (2017)
15. Khan, J.Y., Khondaker, M., Islam, T., Iqbal, A., Afroz, S.: A benchmark study on machine learning methods for fake news detection. arXiv (2019)
16. Loper, E., Bird, S.: NLTK: the natural language toolkit. arXiv preprint cs/0205028 (2002)
17. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed representations of words and phrases and their compositionality. arXiv preprint [arXiv:1310.4546](https://arxiv.org/abs/1310.4546) (2013)
18. Nakamura, K., Levy, S., Wang, W.Y.: r/Fakeddit: a new multimodal benchmark dataset for fine-grained fake news detection. arXiv preprint [arXiv:1911.03854](https://arxiv.org/abs/1911.03854) (2019)
19. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011)
20. Quandt, T., Frischlich, L., Boberg, S., Schatto-Eckrodt, T.: Fake news. *Int. Encyclopedia Journal. Stud.* 1–6 (2019)
21. Rashkin, H., Choi, E., Jang, J.Y., Volkova, S., Choi, Y.: Truth of varying shades: analyzing language in fake news and political fact-checking. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2931–2937 (2017)
22. Shu, K., Liu, H.: Detecting fake news on social media, vol. 11, pp. 1–129. Morgan & Claypool Publishers (2019)
23. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: a data mining perspective. *ACM SIGKDD Explor. Newslett.* **19**, 22–36 (2017)
24. Shu, K., Wang, S., Liu, H.: Understanding user profiles on social media for fake news detection. In: 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 430–435. IEEE (2018)
25. Shu, K., Wang, S., Liu, H.: Beyond news contents: the role of social context for fake news detection. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, pp. 312–320 (2019)
26. Shu, K., Zhou, X., Wang, S., Zafarani, R., Liu, H.: The role of user profiles for fake news detection. In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 436–439 (2019)
27. Sitaula, N., Mohan, C.K., Grygiel, J., Zhou, X., Zafarani, R.: Credibility-based fake news detection. In: Shu, K., Wang, S., Lee, D., Liu, H. (eds.) *Disinformation, Misinformation, and Fake News in Social Media*. LNSN, pp. 163–182. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-42699-6\\_9](https://doi.org/10.1007/978-3-030-42699-6_9)
28. Thota, A., Tilak, P., Ahluwalia, S., Lohia, N.: Fake news detection: a deep learning approach. *SMU Data Sci. Rev.* **1**, 10 (2018)

29. Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G.S., On, B.W.: Fake news stance detection using deep learning architecture (CNN-LSTM). *IEEE Access* **8**, 156695–156706 (2020)
30. Wang, W.Y.: “Liar, liar pants on fire”: a new benchmark dataset for fake news detection. arXiv preprint [arXiv:1705.00648](https://arxiv.org/abs/1705.00648) (2017)
31. Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., Liu, H.: Unsupervised fake news detection on social media: a generative approach. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 5644–5651 (2019)