Fake News Detection Using Machine Learning

Wilson Chuah wchuah@ucsd.edu
Taiyou Chen tac001@ucsd.edu
Xin Cui cxin@ucsd.edu
Yuying Liu yul061@ucsd.edu

Abstract

Recently, the advancement of technologies and the low expense nature of the internet have made the individual's life easy. Additionally, information sharing through these platforms has increased and never been seen before in human history. These social media platforms allow people to share information without any checkpoint or obstacles. Therefore, consumers on these platforms are sharing and creating information that may not be relevant to any reality. This has led to the dissemination of wrong information. There is a need to detect fake news before it spread. This is a difficult task to detect the truthfulness of an article. Even an expert needs to explore the article from different aspects before giving an opinion on whether the news is trustworthy or not. In this paper, we have constructed a model that automatically detects discriminative features from the data and classify either the news content is real or fake. The fake_or_real Kaggle data was used to train the model. Additionally, we have also compared the results obtained by the proposed model with baseline models. The proposed system i.e. SVM with bigram TF-IDF outperformed with an accuracy of 93% using test data. Moreover, the trained model was also evaluated using another dataset namely COVID-19. This work will serve as the baseline model for researchers in the domain of automatic fake news classification.

1 Introduction

In recent years, mobile technologies have enhanced the rapid use of social media platforms including Whats-App, Facebook, Instagram, Twitter, etc. A large amount of data has been producing in a single day on these social media platforms (Ghani et al., 2019). Besides, this online created data majorly news-based (Zhou and Zafarani, 2020). These platforms are the major source of information sharing and communication among users. People rely on these platforms to get updates due to ease of use, fast data transformation, and a less expensive nature. These characteristics enable the user to circulate the data online, which may lead to an increase in the rapid growth of fake news.

This is falsified information that proliferated with the wrong intention. New York Times defines fake news as "made-up stories are written to deceive". It is difficult to identify genuine news from fake news. Because the creators publish the fake news content using the format of real news content (Pan et al., 2018). Therefore, there is a need to build the AI-based robust model to detect fake news (Pan et al., 2018; Shu et al., 2017). Recently, researchers are paying more attention to fake news classification. In this study, we have used a machine learning model for fake news classification. The main goal of this article is to increase the performance accuracy in the given domain.

The rest of the article is arranged as Section 2 highlights the exiting works done by different researchers in the domain of fake news classification. Section 3 provides detailed information about the dataset used in this study. Section 4 explains the steps of the proposed methodology. However, Section 5 explains the experiments that would be served as baseline models. Section 6 presents the experimental results. Lastly, Section 7 concluded the study and provides possible directions for the future in the domain of fake news classification.

2 Related Works

Researchers used different approaches to find solutions to mitigate the adverse effects of fake news dissemination. Fake news detection is a challenging task. But, due to advancements in the field of artificial intelligence, efficient algorithms are available that can be used to detect intentionally created fake news content. Timely detected misinformation

would help to make efficient decisions to avoid the worst effect of misinformation propagation on an individual's life as well as on society. This section covers the work done by different researchers in the field of fake news classification.

(Ott et al., 2011) suggested to use word count and parts of speech tags as main features to investigate spam reviews. The authors obtained 90% accuracy using opinion spam dataset. (Feng et al., 2012) used a novel method for detecting the deceptive content. They have used syntactic stylometry approach with combination of context free grammar rules. They performed experiments using hotel review dataset and obtained 91.2% accuracy. (Pérez-Rosas et al., 2017) have discussed that there is linguistic difference in legitimate and fake news content. Additionally, the authors have compared the manual and automatic detection of fake news. They used a celebrity-news dataset for their experiments and obtained 74% accuracy.

(Ahmed et al., 2017) trained two different machine learning algorithms for fake news classification i.e. linear support vector machine and logistic regression. They used TF-IDF as feature representations. The trained models outperformed with accuracy of 89% and 92% respectively. Additionally, in (Yang et al., 2018), convolutional neural networks was used for the classification of fake news content. In their investigations, they obtained an accuracy of 92%. (Ruchansky et al., 2017) discussed that the important factor for fake news classification is user relationship. The authors obtained an accuracy of 89.2% using a hybrid approach. (Ghanem et al., 2018) detected fake news by using n-gram and word embedding. They obtained an accuracy of 48.8% in their experiments. However,(Singh et al., 2017) have used content as well as context-based approaches for the classification of fake news. They used Word Count and Linguistic Analysis based approach. In their investigations, they obtained 87% accuracy by machine learning classifier i.e., SVM.

3 Fake or Real Dataset

Now days, the news consumption through social media have been increased due to low expenses of mobile or internet technologies. For example, more than 60% of U.S adults used social media to get news updates in 2016 (Pan et al., 2018). Additionally, recent studies showed that the proliferation of news through social media is higher than other traditional news platforms. Therefore, the social media platforms are frequently used to get personal, financial, and political benefits (Allcott and Gentzkow, 2017; Klein and Wueller, 2017). Therefore, people are using these platforms with wrong intentions. This may lead to spreading of fake news. The fake news may have worst effects on individuals as well as on whole society. Like, people accept the wrong or fake beliefs. Secondly, the trust on news media decreased due to dissemination of wrong information. Lastly, the response of people on real news will be changed (Paul and Matthews, 2016; Nyhan and Reifler, 2010). Therefore, it is important to detect the news on time before their spreading. With advancement of artificial intelligence, the systems can be built to detect whether the news is fake or real. There are several datasets available that can be used for fake news classification. In this study, the publicly available kaggle dataset has been used for fake news classification i.e., fake_or_real. The dataset contains four attribute, unnamed attribute, news title, text and its relevant label. The sample of the given dataset is shown in Figure 1. Additionally, Table 1 shows the basic information about the dataset. As shown here, there are two main classes in the dataset i.e whether news are fake or real. The dataset contains 6335 number of instances. However, Figure 2 shows that the given dataset is balanced. The 50.1% of instances belong to real class and remaining 49.9% of instances are related to fake class.

	Unnamed: 0	title	text	label
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fello...	FAKE
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg Linkedin Reddit Stumbleu...	FAKE
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	REAL
3	10142	Bernie supporters on Twitter erupt in anger ag...	— Kaydee King (@KaydeeKing) November 9, 2016 T...	FAKE
4	875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	REAL

Figure 1: Dataset Sample

	Description
Dataset Name	fake_or_real
Description	Data related to news
Source	Kaggle
Labels	Fake & Real
Data-points	6335
Class-wise Instances	Fake: 3164 Real: 3171

Table 1: Dataset Details.

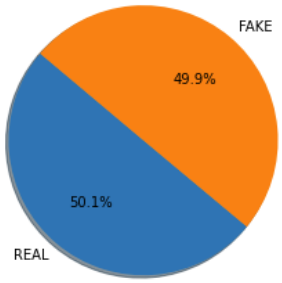

Figure 2: Class wise Instances Distribution

4 Methodology

This section depicts the steps followed to build the model for fake news classification. Figure 3 shows that there are six steps to classify the news content whether it is fake or real namely collection of dataset, preprocessing and data cleaning, feature extraction, splitting of data into train test, fake news model construction and evaluation of constructed model. Additionally, the trained model was also evaluated using user queries or data having no labels. The subsequent sections discuss these steps in detail.

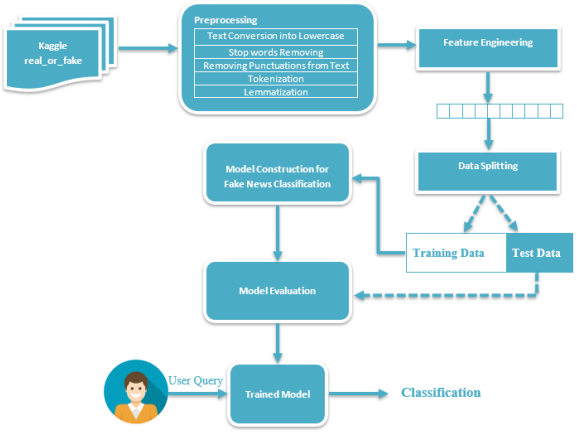

Figure 3: Proposed Methodology for Fake News Classification

4.1 Collection of Dataset

In section 1, it is discussed that the goal of this article is to identify whether the news content is fake or not. The given problem is related to text classification. Thus, the labeled dataset is required to train the classification model. In this research article, the publicly available dataset is used named real_or_fake. As the name shows that the dataset has been labeled into two classes i.e. fake and real. More details of the given dataset have been already discussed in section 3.

4.2 Data Preprocessing

Figure 1 shows that the collected dataset is unclean. To get better classification results, the past literature suggested to clean and preprocess data before model construction. Therefore, in this work, the data has been cleaned using the steps shown in Figure 4. As shown here, firstly the news data has been converted into lower cases. After this, the non-informative features like punctuation, white spaces, stop words, hashtags, and special characters (i.e. @, (,), -, etc.) have been removed. The next step in data preprocessing is tokenization. Tokenization is a method to convert every news into words or tokens. Finally, lemmatization was used. Lemmatization is a technique used to convert words into their root forms like becomes into become, exchanges into exchange, and so on.

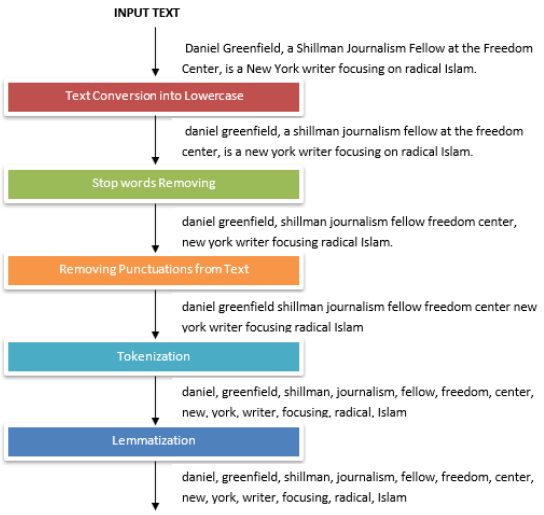

Figure 4: Data Preprocessing

4.3 Feature Engineering

In machine learning, classification algorithms do not understand the raw data. Therefore, to learn the classification rules, there is a need to convert the raw text into numeric forms. In this study, bigram with TF-IDF has been used to extract the features from the raw text and represent these features as numeric vectors.

4.4 Splitting of Data

As shown in Figure 3, the next step after preprocessing is the data splitting. In this study, the cleaned

and preprocessed data was splitted using the pareto principle. According to the principle "80% of effects come from 20% of causes" (Dunford et al., 2014). Therefore, to train the model for news classification, 80% of real_or_fake data was. The constructed model was evaluated using remaining 20% of data. Splinted data for test and training is shown in Table 2.

Class	Instances	Train Data	Test Data
Real	3171	2532	639
Fake	3164	2536	628
Total	**6335**	**5068**	**1267**

Table 2: Instances Distribution According to Pareto Principle

4.5 Model Construction: News Classification

Different machine learning algorithms are used for news classification like a random forest (RF), logistic regression (LR), Naiye Base (NB), Support Vector Machine (SVM), and so on. The SVM classifier uses a threshold value to divide the data regardless of the total features in the data. This means that the SVM classifier considers each feature independently, calculates the probability of each class accordingly, and predicts the label based on the highest calculated probability (Mujtaba et al., 2018). Therefore, In this study, the SVM algorithm was used to train the model for news classification.

4.6 Model Evaluation

For text classification, several performance measures are used to evaluate how the trained model is performing on unseen data. The different performance evaluators are F-measure score, accuracy, precision, recall, etc. These are commonly used performance measures for classification problems. But, it is suggested to use accuracy measures when the target class in the data is well balanced. As Figure 2 shows that the given dataset is approximately equally divided. Therefore, in this study, the accuracy measure has been used to evaluate the news classification model. Additionally, equation 1 describes the accuracy. As shown here, the accuracy is correctly classified instances with respect to the total number of datapoints in the given data.

$$Accuracy = \frac{(\mathrm{TN} + \mathrm{TP})}{\mathrm{TN} + \mathrm{FN} + \mathrm{TP} + \mathrm{FP}} \quad (1)$$

5 Baseline Models

In this study, different machine learning algorithms were used to train a model for news classification that would be served as baseline models. Table 3 shows these machine learning models. However, the other experimental settings for all these baseline models were the same as the proposed model including preprocessing steps, train test split. The details of these baseline models are given in subsequent sections.

#	ML Model	Feature Selection
1	LR	Bigram with TF-IDF
2	MNB	Bigram with TF-IDF
3	RF	Bigram with TF-IDF
4	Ensemble	Bigram with TF-IDF

Table 3: Machine Learning Baseline Models

5.1 Logistic Regression

In text classification, LR is the most commonly used algorithm. It uses a sigmoid or S shaped logistic function to show the relationship between categorical dependent variable and set of independent variables. The algorithm predicts the class of dependent categorical variable. Therefore, the outcome must be discrete form i.e. 0 or 1, yes or no, false or tru, etc. But due to probabilistic based nature, it returns the values between 0 and 1 (Dreiseitl and Ohno-Machado, 2002).

5.2 Multinomial Naïve Base

This machine-learning algorithm uses the "Bayes theorem" for class prediction. According to the term Naïve, the features in the data are mutually independent. This means that the probability of the presence of one feature does not affect by the occurrence of another feature. Therefore, this algorithm can be used as outperforming alternative for small datasets (Lewis, 1998).

5.3 Random Forest

Random forest is supervised machine learning algorithm. It builds the forest. This means that it is an ensemble typed classifier that contains multiple decision trees. The instances are classified based on the votes by different decision trees (Xu et al., 2012).

5.4 Ensemble

In machine learning and statistics, multiple learning algorithms can be the ensemble for constructing a predictive model. Past literature shows that the ensemble classifier shows better performance than individual classifier alone (Hakak et al., 2021). In this baseline model, the three machine learning models namely, LR, RF, and SVM were combined to make the ensemble model.

6 Results and Discussion

In this section, the obtained results by the proposed model have been discussed. Additionally,this section also compared the performance of the proposed model with baseline models. The detailed discussion and results are given in subsequent sections.

6.1 Proposed Model: Results Using Test Data

For performance evaluation of proposed model, 20% of data was used i.e. 1267 instances. The results obtained by proposed model using test are shown in Table 4.

F-Measure	0.93
Recall	0.93
Precision	0.93
Accuracy	93%

Table 4: Experimental Results by Proposed Model Using Test Data

Additionally, Figure 5 shows the confusion matrix of the proposed model using unseen test data. The class wise performance can also be evaluated using the given confusion matrix. As shown here, 1178 out of 1267 instances were correctly classified i.e. 590 as fake and 588 as real. However, 38 and 51 instances were miss classified as real and fake respectively.

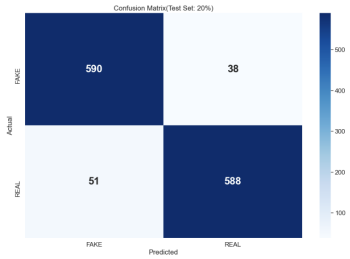

Figure 5: Confusion Matrix of Proposed Model for Test Data

6.2 Proposed Model: Results Using COVID-19 Data

This section shows the results obtained by proposed model using dataset named COVID-19. The given dataset contains news related to COVID. There are 6787 number of instances. The dataset was cleaned by applying the all preprocessing steps. Additionally, bigram with TF-IDF was used to extract the features from the dataset. After this, the extracted features were given to proposed model to predict the class of given news. The predicted results by proposed model are shown in Figure 6.As shown here, 38.9% of 6787 news were predicted as fake and remaining 61.1% news were classified as real i.e. 2640 and 4148 respectively.

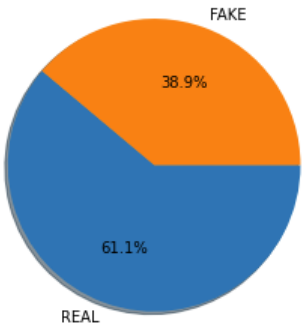

Figure 6: Results by Proposed Model Using COVID-19 Data

6.3 Proposed Model Vs Baseline Models

This section compares the performance of the proposed model with baseline models. The recall (R), precision (P), F-measure score (F), and accuracy (A) by baseline experiments are shown in Table 5. However, Figure 7 depicts the baseline results along with the proposed model results. As shown here, the proposed model outperformed with an accuracy of 93% as compared to the baseline models. It also shows that the results of logistic regression are somewhat lower than the proposed model but higher than the other two baseline models i.e., RFC and MNB. The possible reason behind its lower performance is that there is a probability of having nonlinear data in the fake_or_real dataset. And LR algorithm is linear on the decision surface and cannot handle the nonlinear data efficiently (Eftekhar et al., 2005). The reason behind the low performance of RF might due to the insufficient informative features in the given dataset (Xu et al., 2012). The performance by the MNB classifier

was lowest as compared to other baseline models. As already discussed in section 5.2, the MNB algorithm does not consider the conditional dependence among features, all features are mutually independent of one another. There is the possibility that the fake_or_real data consists of highly dependent features. But these features are voted independently by the model which overinflated the importance of these features and leads to lower performance (Lewis, 1998). Moreover, Table 6 shows the results obtained by different experiments by researchers using same dataset i.e. real_or_fake. As shown here, the performance of the proposed model is also higher than the exiting works.

Baseline	A	P	R	F
LR	92%	0.92	0.92	0.92
MNB	76%	0.83	0.76	0.74
RF	87%	0.87	0.87	0.87

Table 5: Performance Evaluation of Baseline Experiments

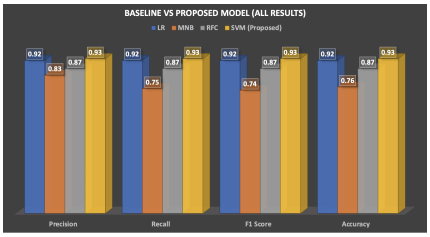

Figure 7: Proposed Model Vs Baseline Models

6.4 Proposed Model Vs Deep Learning Models

This section shows the results obtained with different deep learning models for fake news identification on the same dataset which we used for the baseline as well as the proposed models. Deep learning is a recent advancement in the field of machine learning. It uses the deep neural network architectures to perform several tasks including text classification, information retrieval, sentiment analysis, etc. The major advantage of the deep learning over machine learning is the automatic feature extraction, but it has a major drawback over machine

Authors	Accuracy (%)
(Yang et al., 2018)	92
(Ahmed et al., 2017)	92
(Ruchansky et al., 2017)	89.2
(Ahmed et al., 2017)	89
(Singh et al., 2017)	87
(Ghanem et al., 2018)	48.8

Table 6: Existing Research Using real_or_fake Dataset

learning which is that deep learning requires huge amount of data in order to learn the patterns or to extract the features (Pouyanfar et al., 2018). In this work, we used different deep learning models to identify fake news. Literature suggests that the sequential models in deep learning performs quite well for textual data. So, in this study we used Recurrent Neural Networks (RNN) and its variants. Table 7 shows the list of models which we used in this study. The preprocessing steps, train/test split and other configurations were kept as same as in machine learning experiments.

Model	Reference
Recurrent Neural Network (RNN)	(Hu et al., 2021)
Long-Short Term Memory (LSTM) Network	(Qi et al., 2021)
Bidirectional LSTM	(Fu et al., 2021)

Table 7: Deep Learning Models used in this study

Below section, shows the overall results obtained by the deep learning models mentioned above.

6.5 Results (Deep Learning Models)

This section shows the results obtained by deep learning models for the classification of fake news. All the experimental settings, preprocessing steps and train/test split ratio were kept as same as machine learning models experiments. Table 8 this shows the recall (R), precision (P), F-measure score (F), and accuracy (A) of all the three deep learning models mentioned in table 7.

As we can see that the overall results of all the deep learning models are less than the proposed

DL Model	A	P	R	F
RNN	51%	0.26	0.51	0.35
LSTM	89%	0.89	0.89	0.89
Bidirectional LSTM	90%	0.90	0.90	0.90

Table 8: Performance Evaluation of Deep Learning Models

model. The proposed model obtained the accuracy of 93% but the highest accuracy obtained by the deep learning models is 90%. The main reason for this less accuracy is because deep learning models require huge amount of data to learn the patterns and perform classification. In this study, the dataset we have is not much larger thats why the accuracy of deep learning models is less than the machine learning.

7 Conclusion

It is a time-consuming task to detect the truthfulness of news content manually. Besides, to identify the anomalies in the news content, in-depth knowledge and expertise are required. In this study, we have trained a machine learning model that automatically classifies the problem of fake news. The model was constructed using the dataset available on the Kaggle platform. The constructed model extracts the features from the given data and can differentiate the real news content from fake news content. The trained model not only outperformed on test data with an accuracy of 93%, but it also performed well on the covid-19 news dataset. The results obtained by the proposed model are higher than the other machine learning baseline models including multinomial naïve Bayes, logistic regression, random forest classifier, and other existing studies using the same dataset. This research has limitations like it can only detect fake news in form of text. In the future, a fake news classification model would be constructed that can detect fake news from images or videos. Additionally, deep learning models with pre-trained word embeddings can also be implemented to get better results in the domain of fake news.

References

Hadeer Ahmed, Issa Traore, and Sherif Saad. 2017. 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.

Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36.

Stephan Dreiseitl and Lucila Ohno-Machado. 2002. Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics*, 35(5-6):352–359.

Rosie Dunford, Quanrong Su, and Ekraj Tamang. 2014. The pareto principle.

Behzad Eftekhar, Kazem Mohammad, Hassan Eftekhar Ardebili, Mohammad Ghodsi, and Ebrahim Ketabchi. 2005. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. *BMC medical informatics and decision making*, 5(1):1–8.

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

Yujie Fu, Jian Liao, Yang Li, Suge Wang, Deyu Li, and Xiaoli Li. 2021. Multiple perspective attention based on double bilstm for aspect and sentiment pair extract. *Neurocomputing*, 438:302–311.

Bilal Ghanem, Paolo Rosso, and Francisco Rangel. 2018. Stance detection in fake news a combined feature representation. In *Proceedings of the first workshop on fact extraction and VERification (FEVER)*, pages 66–71.

Norjihan Abdul Ghani, Suraya Hamid, Ibrahim Abaker Targio Hashem, and Ejaz Ahmed. 2019. Social media big data analytics: A survey. *Computers in Human Behavior*, 101:417–428.

Saqib Hakak, Mamoun Alazab, Suleman Khan, Thippa Reddy Gadekallu, Praveen Kumar Reddy Maddikunta, and Wazir Zada Khan. 2021. An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generation Computer Systems*, 117:47–58.

Zexin Hu, Yiqi Zhao, and Matloob Khushi. 2021. A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, 4(1):9.

David Klein and Joshua Wueller. 2017. Fake news: A legal perspective. *Journal of Internet Law (Apr. 2017)*.

David D Lewis. 1998. Naive (bayes) at forty: The independence assumption in information retrieval. In *European conference on machine learning*, pages 4–15. Springer.

Ghulam Mujtaba, Liyana Shuib, Ram Gopal Raj, Retnagowri Rajandram, and Khairunisa Shaikh. 2018. Prediction of cause of death from forensic autopsy reports using text classification techniques: A comparative study. *Journal of forensic and legal medicine*, 57:41–50.

Brendan Nyhan and Jason Reifler. 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2):303–330.

Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T Hancock. 2011. Finding deceptive opinion spam by any stretch of the imagination. *arXiv preprint arXiv:1107.4557*.

Jeff Z Pan, Siyana Pavlova, Chenxi Li, Ningxi Li, Yangmei Li, and Jinshuo Liu. 2018. Content based fake news detection using knowledge graphs. In *International semantic web conference*, pages 669–683. Springer.

Christopher Paul and Miriam Matthews. 2016. The russian.

Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2017. Automatic detection of fake news. *arXiv preprint arXiv:1708.07104*.

Samira Pouyanfar, Saad Sadiq, Yilin Yan, Haiman Tian, Yudong Tao, Maria Presa Reyes, Mei-Ling Shyu, Shu-Ching Chen, and SS Iyengar. 2018. A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)*, 51(5):1–36.

Ling Qi, Matloob Khushi, and Josiah Poon. 2021. Event-driven lstm for forex price prediction. *arXiv preprint arXiv:2102.01499*.

Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. 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.

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

Susheela Singh, Altaf Hossain, Isaac Maddow-Zimet, Michael Vlassoff, Hadayeat Ullah Bhuiyan, and Meghan Ingerick. 2017. The incidence of menstrual regulation procedures and abortion in bangladesh, 2014. *International perspectives on sexual and reproductive health*, 43(1):1–11.

Baoxun Xu, Xiufeng Guo, Yunming Ye, and Jiefeng Cheng. 2012. An improved random forest classifier for text categorization. *JCP*, 7(12):2913–2920.

Yang Yang, Lei Zheng, Jiawei Zhang, Qingcai Cui, Zhoujun Li, and Philip S Yu. 2018. Ti-cnn: Convolutional neural networks for fake news detection. *arXiv preprint arXiv:1806.00749*.

Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40.