DOI: 10.15849/IJASCA.220328.04

Int. J. Advance Soft Compu. Appl, Vol. 14, No. 1, March 2022
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Empirical Evaluation of Machine Learning Classification Algorithms for Detecting COVID-19 Fake News

Hiba Alsaidi and Wael Etaiwi

King Talal School of Business Technology, Princess Sumaya University for Technology, Amman, Jordan
hib20198090@std.psut.edu.jo, w.etaiwi@psut.edu.jo

Abstract

Humans have been fighting the Covid19 pandemic since it started, not just to protect their wellbeing but also to counteract the news and rumors that have been spreading about it. Rumors and false allegations can be almost as dangerous as the virus, as they affect people's mental health and increase their stress levels. To address this problem, several machine learning techniques could be used to detect fake news. In this paper, four different machine learning algorithms are compared according to their ability to detect fake news, including Naive Bayes, Decision Tree, Support Vector Machines, and Logistic Regression. A dataset of annotated news is used in the experiments. The experimental results show that Naïve Bayes outperforms other algorithms in terms of accuracy, precision, recall, and F1 score.

Keywords: *COVID-19, Machine Learning, Fake news detection.*

1. Introduction

Because of the increasing influx of information around the world, fake news has become a major problem. While social media and blogs are beneficial in several respects, such as sharing important information and facilitating contact, they eventually became a burden, a deceptive source to readers, and a source of negative reactions as a result of the dissemination of fake news and reviews. People would be unable to resist extracting knowledge from it, whether true or false conclusions. For example, fake

Received 6 November 2021; Accepted 23 January 2022

news has been pivotal in the election of President Trump [1], and 40% of the most frequently shared medical links contained fake news or information [2]. As a result, the need for data mining techniques that can detect fake news has arisen, and these techniques have proven to be extremely useful in preventing fake news from being transmitted over the internet [3].

COVID-19 is a coronavirus-related infection caused by a newly discovered coronavirus. Coronavirus infection causes mild to moderate aerobic sickness, which normally resolves without the need for special care [4]. People with diabetes, cardiovascular disease, chronic respiratory disorder, the elderly, and cancer patients, on the other hand, are more likely to develop severe illnesses. Saliva droplets, coughs, and sneezes transmit the coronavirus. People should have a clear understanding of the COVID-19 virus, its causes, and how it spreads to stop disease transmission. People should wash their hands, wear a face mask, rub their hands with alcohol, avoid rubbing their noses, and cough into a flexed elbow to protect themselves and others from infection. In addition to safety precautions, it is important that people recognize the difference between fake and real news about the disease, as being misled by fake news can hurt people's attitudes and psyche [5].

Data scientists are still trying to figure out the best ways to analyze data to derive valuable information from it [6]. Four different machine learning algorithms are used in this paper research to examine the ability of machine learning in detecting COVID-19 fake news, including Decision tree (DT), Naïve Bayes (NB), Logistic Regression (LR), and Support Vector Machines (SVM).

DT is a divide-and-conquer classification technique that is used to extract patterns and discover features from a dataset. Both exploratory data analysis and predictive analytics applications have made extensive use of decision trees [7]. Most decision-tree classifiers (such as C4.5) function in two stages: decision tree construction and decision tree pruning [8]. The decision tree model is constructed in the first step by recursively splitting the training data set using a locally optimal criterion. The leaves and branches that are responsible for classification are pruned in the second stage. NB is a supervised learning algorithm that uses the Bayes theorem and naive assumptions about strong independence among features to create a probabilistic classifier [9]. It is simple to construct because features are normally distributed, useful for very large data sets, and considered quick in training. Sentiment Analysis and Spam Filtering are two typical applications. The algorithm of LR is used in classification and regression tasks [10]. It associates independent variables, either continuous or categorical, with one dependent variable, and it models the likelihood of a class, such as a pass or fail, false or true, alive or dead, or yes or no. The response variable is binary and can only be allocated to one of two classes. It is commonly used to predict categorical variables from dependent variables. For classification and regression problems, SVM is a supervised learning algorithm that divides n-dimensional space into sections using the best decision boundary (called a hyperplane) [11].

Many open-source software, such as KNIME[1], RapidMiner[2], and WEKA[3], are designed for data mining purposes. They are used to analyze data using a variety of pre-implemented algorithms, methods, and techniques, such as neural networks, decision trees, statistical analyzers, regression analyzers, data visualization, text mining, and predictive analytics. In this paper, the KNIME Analytics Platform is used in the experiments. In this paper, empirical experiments are conducted to compare the performance of four different classification algorithms: decision tree, Naïve Bayes, LR, and Support Vector Machines. The algorithms are compared according to their ability to detect Covid-19 fake news. The performance analysis is evaluated using the confusion matrix in terms of precision, recall, and f-score.

The remaining of this paper is organized as follows: The literature review is discussed in section 2. In section 3, the methodology, the dataset, and the experimental results are discussed. Finally, in section 4, the conclusion is provided.

2. Literature review:

Detecting spam/fake reviews and news has become a well-studied research topic for academics, like social media, websites, and blogs constantly pump out news and articles, necessitating the need to know what is real and what is false.

Allcott and Bing Liu et all investigated review spam detection over two million reviewers and five million reviews derived from the Amazon website [12, 1]. The investigation findings highlighted the widespread spam activities, based on duplications in findings and classifications (spam and non-spam). LR was used to learn the predictive model, and the results showed the model's effectiveness.

Tarek et al proposed an ensemble approach for spam detection in Arabic reviews [13]. The proposed model combines text mining and data mining as a single mining classification method. In the experiments, the sampling method is used to overcome

[1] https://www.knime.com/
[2] https://rapidminer.com/
[3] https://www.cs.waikato.ac.nz/ml/weka/

class imbalance issues. And the dataset was gathered from various sources such as Booking, TripAdvisor, and Agoda datasets. With an F-measure of 99.59%, the experimental results showed that the proposed model was effective at detecting Arabic spam reviews.

Using three different datasets from news reviews and articles, Hadeer et al. [14] developed an n-gram model to automatically expose fake contents by comparing term frequency inverted document frequency (TF-IDF) with term frequency (TF) using six machine learning classification algorithms: Stochastic Gradient Descent, SVM, linear SVM, DT, LR, and K-nearest neighbor. The experimental evaluation showed very encouraging and improved performances compared to the state-of-the-art methods in terms of accuracy.

Zongru et al. [15] detected fake news from online articles by comparing three machine learning techniques: Recurrent Neural Networks (RNN), random forest, and NB, using an accessible dataset from the Kaggle[4] website for detecting fake news. Different semantic and statistical features such as TF-IDF, TF, quadgram, trigram, bigrams were used in the experiments. The experimental results showed that the random forest algorithm with bigram features had the highest accuracy of 95.66%. On the other hand, Barushka and Hajek [16] used two hotel reviews datasets to build a vector model from *n*-grams and skip-gram word embedding methods. By using a feed-forward neural network to classify review spam correctly, the experimental results showed that the proposed model outperforms other review spam detection algorithms in terms of accuracy and Receiver Operating Characteristics (ROC).

A new machine learning classification model to classify suspicious and non-suspicious 7000 Bengali articles was proposed by Hoque et al. [17]. The training set contains 5600 documents, and the testing set contains 1400 documents. To test the performance of the proposed model, human-based and machine learning-based techniques were compared, and the proposed model overcomes other models and achieved an accuracy of 84.57%.

To detect hate tweets in an Arabic dataset, Al-Khalifa et al. [18] used a combination of neural network algorithms, including RNN and Convolutional Neural Networks (CNN). A dataset with 9,316 hate tweets was created and divided into three categories: abusive, hateful, and normal. Then, to detect Arabic hate tweets, four performance models were analyzed and compared including Bidirectional Encoder Representations from Transformers (BERT), CNN, Gated recurrent units (GRU), and

[4] https:\\www.kaggle.com\

CNN+GRU. With an Area Under the Receiver Operating Characteristics (AUROC) of 89% and an F1-score of 79%, the findings indicated that CNN outperformed other models.

Alami and El Beqqali [19] developed an automated system to identify suspicious social media profiles by calculating similarity distance (Normalized Compression Distance) to detect suspicious posts, by comparing social media posts against a predefined database of suspicious posts, using the similarity method in text analysis. The findings take advantage of the significance of similarity distance.

Andreas et al. [20] presented a unique data set for rumor debunking extracted from a digital journalism project. The dataset included 2,595 news articles and 300 rumored claims classified as true, false, or unverified. Individual articles were summarized and labeled, emergent recognized the claim, and each was assigned to a class. The proposed model used an LR classifier to identify the article headline to the claim after presenting the data. Then, the Excitement Open Platform textual entailment classifier is compared to the proposed classifier. The experimental results showed that the proposed classifier outperformed the Excitement Open Platform by 73%.

In addition to the aforementioned research, scientists are currently attempting to develop artificial intelligence methods to deal with fake news related to the COVID-19 pandemic in 2020. Gautam et al. developed a COVID-19 dataset with over 5,000 fact-checked articles from over ninety different fact-checking websites and annotated these articles into 11 different categories [21]. Then they created a classifier model to detect Fake Covid news, with an F1 score of 0.76. Rutvik et al. developed an annotated dataset of COVID-19-related Hindi and Bengali tweets and proposed a BERT-based model to detect fake COVID-19 tweets [22]. In fake tweet detection, the proposed method received an F-Score of 89%. Furthermore, Jindal used a dataset of 10,700 fake and real news articles related to COVID-19 to detect fake news [23]. This data was gathered from a variety of social media sites and websites. The performance of three machine learning algorithms was compared on the constructed dataset. The algorithms include DT, Gradient Boost, SVM, and LR. The experimental results showed that SVM outperformed the other algorithms with an F1-score of 93.46%, followed by LR with an F1-score of 92.75%.

3. Methodology and Evaluation

Figure 1 illustrates the proposed methodology's major phases. Preprocessing, feature extraction, applying machine learning algorithms, extracting results, and performing a comparative assessment are all steps in the proposed methodology.

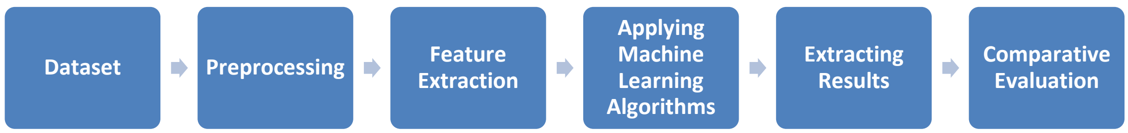

Figure 1: The Proposed Methodology Steps

The preprocessing phase aims to cleanse the dataset and prepare it for use in the next upcoming steps. In the preprocessing step, punctuation marks are removed, letter case is normalized, stop words are removed, and stemming is applied. Punctuation marks are deleted and omitted from the entire text because they have no impact on the text classification decision. The letter case normalization step is used to convert all letter cases to small letters to eliminate letter case confusion. Pronouns, linking words, and prepositions are among the stop words that have been removed. Finally, stemming is used to consider stems rather than words to minimize words diversity.

The feature extraction step aims to extract useful features from the given dataset to prepare it for the text classification step. To improve text classification, several types of features are used. A collection of features, such as keywords, IDF and TF frequencies, term c-occurrence counter, document vector, document data extractor, and tag cloud are extracted from the text.

Different machine learning algorithms were used for text classification after preprocessing and feature extraction. The dataset was evaluated with LR, DT, NB, and SVM.

3.1. Dataset

To build and evaluate the targeted model, a dataset of annotated articles is used in the experiments. The dataset contains news articles that were widely published during the coronavirus pandemic and was shared on the Figshare website[5] for binary classification, temporal, and spatial features. After removing duplicates, the data to be

[5] https://figshare.com/

used in building the model consists of 2472 rows and 6 features: (Article title, text, country, source, date, and target class).

The target class indicates whether the article is correct and genuine about Covid-19, or whether it is a piece of fake news that makes claims and speculations about Covid-19. The dataset is imbalanced by class, with 63% of fake articles and 37% of the correct articles.

Correct Articles Fake Articles The Whole Dataset

Figure 2: Most frequent word in the dataset

Table 1: Dataset Statistics

Articles	The average length of sentences	The Average Number of words	The Number of Unique words
Correct articles	154.94	4.9	3368
Fake articles	471.62	5.5	5294
Correct & Fake articles	626.56	10.4	8662

Table 1 shows that fake articles have longer sentences than correct articles based on the average number of words and sentence length. Furthermore, the most ten frequent words in the correct articles after removing punctuation marks, redundant characters, and stop words are: coronavirus, China, health, Facebook, video, Hospital, outbreak, post, covid, people. On the other hand, the most ten frequent words in the fake articles after applying the same preprocessing of correct articles are: coronavirus, China, virus, claim, Facebook, post, covid, vaccine, disease, and people. Finally, the most ten frequent words in the whole dataset are coronavirus, China, health, virus, covid, people, post, media, Facebook, and death. It is noted that there is a considerable overlap of words between fake and correct news, as evidenced by word clouds and frequent words shown in Figure 2.

3.2. Evaluation measures

The main goal of the experiments is to evaluate the ability of different machine learning algorithms to extract fake news. The evaluation metrics include precision, recall, F-score, and accuracy. The precision represents the fraction of the correct decisions to the total number of the given decisions in a particular class. It is calculated as:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

Where *True Positive* is the correct fake news decisions, and *False Positive* is the incorrect fake news decisions. The recall refers to the fraction of the correct decisions that are given by the machine learning method to the total number of news in a particular class. It is calculated as:

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

Where *False Negative* is the incorrect not-fake news decisions. The accuracy measures how close the obtained decisions are to the actual classification. The evaluation process is a binary decision process in which the decision of the machine learning model is correct if and only if it matches the real fake news class in the dataset. It is calculated as:

$$Accuracy = \frac{TruePositive + TrueNegative}{TruePositive + TrueNegative + FalsePositive + FalseNegative}$$

Where *True Negative* is the correct not-fake news decision. Finally, the F1-score is the harmonic mean of precision and recall. It is calculated as:

$$\text{F1} - \text{score} = \frac{2 * Precision * Recall}{Precision + Recall}$$

3.3. Experimental results

Each algorithm is applied ten times after splitting the dataset for training and testing parts. The training dataset consists of 70% of the dataset. While the testing dataset forms the remaining 30% of the dataset. The averages of all experiments are listed in table 2 in terms of Accuracy, precision, recall, and F1 score.

Table 2: Experimental Results

	Accuracy	Precession	Recall	F1 Score
Decision Tree	**78.85**	**81.22**	**82.32**	**81.77**
Logistic Regression	**83.5**	**85.24**	**88.2**	**86.69**
Naïve Bayes	**94.6**	**95.24**	**92.63**	**93.92**
Support Vector Machine	**85.4**	**84.58**	**82.32**	**83.43**

The experiments are conducted using the Knime Analytics Platform. The experimental results indicate that different machine learning algorithms produce different results. In terms of all evaluation metrics, it is obvious that NB outperforms other machine learning algorithms. This indicates that the probability machine learning algorithm is most sufficient for classifying text, especially, when the dataset features are extracted statistically.

More features were required to train and include in the final model for DT algorithms, resulting in better results and a more efficient classifier. DT is also superior for categorical data classification. The total number of features in our dataset is small, which may explain why DT performs poorly when compared to other machine learning algorithms. Even though both algorithms perform well when the training dataset is small, the experimental results show that SVM outperforms LR. This is because SVM is better at handling outliers than LR. The outlier in our dataset refers to exceptionally lengthy or very short news in comparison to other news lengths. NB is performing better when the features are mutually independent, and also works well with small datasets with a small number of features since NB is a generative statistical model.

4. Conclusion

In this paper, a correct-fake Covid19 news dataset was used to classify and detect fake news. The proposed methodology consists of the following steps: preprocessing, feature extraction, applying different machine learning algorithms, and evaluation and comparison. Dataset was run on different machine learning models including Logistic Regression, Decision Tree, Naïve Bayes, and Support Vector Machine. The experimental results showed that Naïve Bayes outperforms other algorithms in terms of accuracy, precision, recall, and F1 score.

References

[1] M. G. Hunt Allcott, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 211-236, 2017.

[2] W. K.-W. A. K. Przemyslaw M.Waszak, "The spread of medical fake news in social media – The pilot quantitative study," *Health Policy and Technology,* vol. 7, no. 2, pp. 115-118, 2018.

[3] M. Y. S. Y. a. M. O. A. Iftikhar Ahmad, "Fake News Detection Using Machine Learning Ensemble Methods," *Complexity,* vol. 2020, 2020.

[4] B. H. M. M. a. P. K. Amri Hammami, "Physical activity and coronavirus disease 2019 (COVID-19): specific recommendations for home-based physical training," *Managing Sport and Leisure,* pp. 1-6, 2020.

[5] L. Bender, "Interim Guidance for COVID-19 Prevention and Control in Schools," *UNICEF,* 2020.

[6] D. J. H. N. M. Adams, Wiley StatsRef: Statistics Reference Online, American Cancer Society, 2015, pp. 1-7.

[7] R. N. L. A. a. S. D. Anthony J.Myles, "An introduction to decision tree modeling," *JOURNAL OF CHEMOMETRICS,* pp. 275-285, 2004.

[8] W. D. a. Z. Zhan, "Building Decision Tree Classifier on Private Data," *Electrical,* vol. 8, 2002.

[9] S. D. a. C. H. Jadhav, "Comparative study of K-NN, naive Bayes and decision tree classification techniques," *International Journal of Science and Research (IJSR),* vol. 5, no. 1, pp. 1842--1845, 2016.

[10] S. Menard, Applied logistic regression analysis, vol. 106, SAGE, 2002.

[11] V. N. Vapnik, "An overview of statistical learning theory," *IEEE transactions on neural networks,* vol. 10, no. 5, pp. 988-999, 1999.

[12] Jindal, Nitin, and Bing Liu, "Review spam detection," in *Proceedings of the 16th international conference on World Wide Web. ACM,* 2007.

[13] Radwa M.K. , Sherine R , Tarek F. , "An ensemble approach for spam detection in Arabic opinion texts," *Journal of King Saud University,* pp. ISSN 1319-1578, 2019.

[14] Ahmed, Hadeer et al., "Detecting opinion spams and fake news using text classification," *Security and Privacy,* vol. 1, no. 1, 2017.

[15] Pranav B., Zongru S., "Fake News Detection with Semantic Features and Text Mining," *hal-02071914f,* 2019.

[16] A. Barushka & P. Hajek, "Detection Using Word Embeddings and Deep Neural Networks," 2019, pp. 340-350.

[17] Sharif, O.; Hoque, M.M et al., "Detecting Suspicious Texts Using Machine Learning Techniques.," *Applied Sciences, section Computing and Artificial Intelligence,* 2020.

[18] Alshalan, R & Al-Khalifa, H. A , "Deep Learning Approach for Automatic Hate Speech Detection in the Saudi Twittersphere," *Applied Sciences,* 2020.

[19] Alami, S. & El Beqqali, O., "Detecting suspicious profiles using text analysis within social media.," *Journal of Theoretical and Applied Information Technology*, 2015.

[20] Vlachos, William F. & Andreas, "Emergent: a novel data-set for stance classification," in *In Proceedings of the 2016 Conference of NAACL*, 2016.

[21] Gautam S. & Durgesh N., "a multilingual cross-domain fact check news dataset for covid-19," *arxiv*, 2020.

[22] Debanjana K, Mohit B, Suranjana S, & Amar P., "No rumours please! a multi-indic-lingual approach for covid fake-tweet detection," *arXiv*, 2020.

[23] Rutvik V., Prathyush P., Sddharth K., & Tyagi S., "Two stage transformer model for covid-19 fake news detection and fact checking," *arxiv*, 2020.

[24] N. a. B. L. Jindal, "Review spam detection," in *Proceedings of the 16th international conference on World Wide Web. ACM,*, 2007.

[25] I. Agarwal. [Online]. Available: https://figshare.com/articles/dataset/COVID19FN_Dataset_csv/12489293/1.

[26] Sofea A. & Izzatdin A., "rorism Detection Based on Sentiment Analysis Using Machine Learning," *Journal of Engineering and Applied Sciences*, 2017.

Notes on contributors

Hiba Alsaidi *received her M.Sc student in Business Analysis at Princess Sumaya University for Technology. She is pursuing research on data mining, machine learning, data and business analysis, e-commerce, e-business, and e-learning.*

Wael Etaiwi *is an assistant professor in the Department of Business Information Technology at Princess Sumaya University for Technology, Jordan. He received his BSc degree in Computer Information Systems from the Hashemite University in 2007, his M.Sc Degree in Computer Science in 2011 from Al Balqaa Applied University, and his Ph.D. in Computer Science from Princess Sumaya University for Technology in 2020. His research interests include, but are not limited to, Artificial intelligence, Data mining, and Natural Language Processing.*