CTD International Journal for Media Studies *Online version ISSN xxxx-xxxx*
Vol. 2 No. 2, 2024: P.1 - 7

Original Article *Available online at www.ctdjms.com*

Application of Supervised Machine Learning Algorithms for Detection of Fake News using Support Vector Machine Classifier

Dong Young Lee[1] and Yuan-Yuan Liu[2]

[1]Information Systems Department, Pukyong National University, Republic of Korea
[2]Harbin University of Commerce, Heilongjiangm, China

Received: 12 July 2024, Revised: 26 September 2024, Accepted: 25 October 2024, Published: 4 November 2024

How to cite this article: Lee, D. Y. & Liu, Y. Y. (2024). Application of Supervised Machine Learning Algorithms for Detection of Fake News using Support Vector Machine Classifier. *CTD International Journal for Media Studies*, 2(2), 1-7

ABSTRACT
The spread of fake news has become a critical issue in the digital age, necessitating the development of robust detection methods. This study explores the application of supervised machine learning algorithms, specifically the Support Vector Machine (SVM) classifier, for the detection of fake news. The study utilizing the publicly available ISOT Fake News dataset from Kaggle. The SVM demonstrated exceptional performance, achieving an overall accuracy of 99%, underscoring its effectiveness in distinguishing between false and true news. The confusion matrix further reveals strong classification results, with 4,542 fake news items correctly identified and only 26 true articles misclassified as fake. Conversely, the model accurately identified 4,264 true news items, with just 20 fake articles incorrectly classified as true. These findings highlight the potential of SVMs in enhancing the reliability of news detection systems and contribute to ongoing efforts to combat misinformation in the media landscape.

Keywords: Fake News, Machine Learning, Support Vector Machine, Digital Age.

1. INTRODUCTION

The proliferation of information in the digital age has revolutionized the way we consume news, but it has also given rise to a significant challenge: the spread of fake news. With social media platforms serving as major conduits for information, the rapid dissemination of misleading content can have serious implications for public opinion, political processes, and societal trust (Lazer et al., 2018). As a result, there is an urgent need for effective mechanisms to detect and mitigate the impact of false information (Pennycook & Rand, 2018).

The consequences of fake news are far-reaching, affecting everything from individual decision-making to the functioning of democratic institutions. The ability to quickly and accurately identify false narratives is crucial for maintaining a well-informed public. In this context, supervised machine learning algorithms have emerged as powerful tools for addressing this challenge, offering the ability to analyze vast datasets and identify patterns indicative of fake news. Among these algorithms, the Support Vector Machine (SVM) classifier stands out for its robustness and effectiveness in high-dimensional spaces (Cortes & Vapnik, 1995). SVMs operate by finding the optimal hyperplane that separates data points into distinct classes, making them particularly well-suited for binary classification tasks such as distinguishing between real and fake news articles.

Recent advancements in natural language processing (NLP) and feature extraction techniques have further enhanced the capability of SVMs to analyze textual data, allowing for a deeper understanding of linguistic features that characterize fake news (Zubiaga et al., 2016). This study aims to explore the application of supervised machine learning algorithms, with a focus on the SVM classifier, for the detection of fake news. By leveraging feature extraction techniques and employing a diverse dataset of news articles, we will evaluate the performance of the SVM in accurately identifying misleading content. Through this research, we seek to contribute to the ongoing efforts to enhance media literacy and promote the integrity of information in an increasingly complex digital landscape.

1.1 Fake News Classification

The categorization of fake news is multifaceted and diverse, encompassing everything from unverified hearsay circulating on social media to deceitful propaganda deliberately spread by its creators. According to references [22–25], fake news is classified into five categories, illustrated in Table 1: (1) deceptive fake news; (2) false information of rumor nature; (3) false comment information; (4) headline party-type fake news; (5) fact-based

recombination of false information.

Table 1: Classification and explanation of Fake News

Fake News Classification	Definition
Deceptive fake news	A false information intended to mislead and deceive the reader. Deceptive fake news is more deceptive and is intended to deliberately mislead readers or cause adverse effects.
False information of rumor nature	Unconfirmed rumors, rumors or anonymous messages, etc.
False comment information	An untrue or misleading comment posted on an online platform, social media, or other interactive platform.
Headline party-type fake news	Edit false headlines eye-catching, the actual content but no reference value of the news
Fact-based recombination of false information	To create misleading or false impressions by reorganizing true facts.

1.2. Research Methods of Fake News Detection

Most of the existing research on fake news detection methods regards fake news detection as a classification task. At present, according to the main features used by the classification model, from the perspective of methods, fake news detection can be divided into three categories: content-based detection methods, social network-based detection methods and knowledge-based detection methods.

1.2.1 Content-Based Detection Method

The content-based fake news detection method aims to extract various semantic features from the news content and detect the authenticity of the news through these features. There are some linguistic differences between fake news and true news, and fake news can be detected by distinguishing the language style of true and fake news texts. Fake news is more subjective than real news.

1.2.2 Detection Method Based on Social Network

The content-based approach can discover the linguistic features of true and fake news. However, sometimes fake news will mislead readers by deliberately imitating the writing techniques of real news. The content-based approach cannot distinguish the feature differences between such fake news and true news.

1.2.3 Knowledge-Based Detection Method

Knowledge-based (KB) fake news detection detects the authenticity of news by verifying fake news and facts, so this is also called fact checking. Fact checking can be divided into two categories: manual verification and automatic verification [6]. The manual method uses domain expert knowledge or the crowdsourcing method. It has high accuracy but low efficiency, which cannot meet the needs of the era of big data. The automatic verification method using natural language processing and machine learning technology has become a hot research field. Fact checking first needs to construct a knowledge base or knowledge graph from the network through knowledge extraction. Then it compares and verifies the fake news with the knowledge base or knowledge graph to judge the authenticity of the news.

1.2.4 Multimodal Fake News Detection

In addition to the detection method based on a single feature source, it can also combine multiple features for fake news detection. In recent years, the data used in fake news detection is no longer limited to text information, and there has been an increasing focus on visual features. Multimodal fake news detection refers to the use of multiple types (such as text, images, etc.) of data to determine whether a news report contains misleading or inaccurate content [36–38]. Cao et al. [39] found that visual content has become an important part of fake news. Fake news often uses unverified visual content (video, images, etc.) to mislead readers and deepen their trust in false information. Pictures, videos and other media information can also be applied to fake news detection.

2. RELATED WORK

Costin BUSIOC et. al. (2020) has used linear regression algorithm to detect fake news. And to propel the model, a dataset of fake and true news has been created. Because in order to train the machine, it must have experience of both types of news. If the machine has news of both the ways, then it will be able to take the right decision. The author has used 65% true news and 35% false news and has tried his based pay machine. Then using this as a basis, it is divided into Trined and Test data sets, whose ratio is 8: 2. The experiment has achived 91%

accuracy.

Jiang et al. (2021) proposes an ensemble method of stacking of logistic regression, decision tree, k-nearest neighbor, random forest, and support vector machine (SVM). All of these approaches achieved an accuracy of over 85%, for verification of real-time generated news. These models find their applications in various systems, including private sector news, public sector news, and government news broadcast channels. The speed of detection is slow in these network models due to high complexity neural networks.

Silva et al (2020) utilized a random forest (RF) classifier by incorporating twenty-three textual features. It applied different feature selection techniques to identify the most significant features. They compared results with benchmark techniques like GBM, XGBoost, and the Ada Boost regression model.

Felber (2021) presented in comparative study of machine learning classifiers which includes support vector machine (SVM), naïve Bayes (NB), random forest (RF), and logistic regression (LR) to detect Fake News. The study evaluated the performance of these classifiers on diverse datasets. SVM model has achieved the highest accuracy on the liar, fake job posting, and fake news datasets.

Bharadwaj and Shao (2019) used recurrent neural networks, a Naïve Bayes classifier, and random forest classifiers on the kaggle.com fake news dataset using six feature extraction techniques: TF, TFIDF, Bigram, Trigram, Quadgram, and GloVe. This team got quality results using bigram features with the 95.66% random forest classifier.

Kaur et al. (2020) proposed a new methodology to detect fake news by using the combined technique of seven machine learning classifiers. Kaliyar (2018) has studied natural language processing using a wide range of machine learning and deep learning classifiers to sort fake news datasets submitted by the Kaggle community. In one of his studies, he applied two feature extraction techniques: TF and TF-IDF. This researcher achieved the best classification accuracy, 98.3%, using the CNN algorithm.

3. MACHINE LEARNING (ML) CLASSIFICATION

Machine Learning (ML) is a class of algorithms that help software systems achieve more accurate results without having to reprogram them directly. Data scientists characterize changes or characteristics that the model needs to analyze and utilize to develop predictions.

3.1 Logistic Regression

Logistic regression is another Supervised Machine learning algorithm which is used for data that is co-dependent on each other, such as heads or tails. It is used to capture the relation between the binary valued data and convert that into a function based on one dependent variable and one or more independent variables.

3.2 Naïve Bayes

This classifier is mainly suitable for contextual data and hence has been used, as it suits the nature of the dataset used. It is built on the principle of Bayes theorem. The multinomial naïve bayes classifier has been used to determine the category of document and make the prediction based on the regularity of words in the file.

3.3 Decision Tree

The decision tree is an important tool that works based on flow chart like structure that is mainly used for classification problems. Each internal node of the decision tree specifies a condition or a “test” on an attribute and the branching is done on the basis of the test conditions and result. Tree based learning algorithms are widely with predictive models using supervised learning methods to establish high accuracy. They are good in mapping non-linear relationships. They solve the classification or regression problems quite well.

3.4 Random Forest (RF)

Random forest (RF) is an advanced form of decision trees (DT) which is also a supervised learning model. RF consists of large number of decision trees working individually to predict an outcome of a class where the final prediction is based on a class that received majority votes. The error rate is low in random forest as compared to other models, due to low correlation among trees (Gregorutti et. al, 2017). Random forest model was trained using different parameters; i.e., different numbers of estimators were used in a grid search to produce the best model that can predict the outcome with high accuracy. There are multiple algorithms to decide a split in a decision tree based on the problem of regression or classification. For the classification problem, we have used the Gini index as a cost function to estimate a split in the dataset. The Gini index is calculated by subtracting the sum of the squared probabilities of each class from one. The mathematical formula to calculate the Gini index (*G*ind) is as follows (Breiman et. al, 1984)

$$G_{\text{ind}} = 1 - \sum_{i=1}^{c} (P_i)^2,$$

3.5 K-nearest neighbor (KNN)

This is an algorithm characterized by simplicity and ease of implementation, derived from supervised machine learning algorithms, which were first used in the early 1970s to solve classification and regression problems. This algorithm is based on the principle of similarity between neighbors. Since the principle of similarity depends on the value of K, cases are classified by the majority of the votes of their neighbors. This occurs because similar cases are close to each other (Alkhatib et al., 2013; Abdulqader et al., 2020).

3.6 Support Vector Machine

Support vector machine (SVM) is another model for binary classification problem and is available in various kernels functions (Cristianini & Shawe-Taylor, 2000). The objective of an SVM model is to estimate a hyperplane (or decision boundary) on the basis of feature set to classify data points (Hofmann et. al, 2008). The dimension of hyperplane varies according to the number of features. As there could be multiple possibilities for a hyperplane to exist in an *N* dimensional space, the task is to identify the plane that separates the data points of two classes with maximum margin. A mathematical representation of the cost function for the SVM model is defined as

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

such that

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

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

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

4. MATERIAL & METHOD

4.1 Datasets

The publicly available dataset in the Kaggle called ISOT Fake News is used to train and test the models developed in this research. It consists of two CSV files, named 'Fake.csv' and 'True.csv'. 'Fake.csv' file holds fake articles scraped from different unreliable websites. These websites are marked as the least trusted websites by PolitiFact and Wikipedia. The articles present in this CSV file contains fake news articles that cover topics such as Government News, Middle-east, US News, Left-news, politics and News. The second CSV file named 'True.csv' holds articles scrapped from a trusted international news firm called Reuters. The articles presented in this dataset falls into the category of Political and world news.

4.2 Data Pre-processing

By the nature of the dataset used, it contains a lot of noise being a natural language. To make the data apt for the algorithms to work on, it is undergoes various computations. Data normalization is a necessary pre-processing step. To begin with clearing of the data, the identification of punctuation marks and stop-words is done followed by their removal. Then the data is tokenized and converted into lower case by calling a function to remove imbalance among them. This process shall shorten the dataset by removing the unnecessary data. To differentiate the fake news and true news in the dataset, a target variable column called label is created and labelling is performed where "0" is labelled for fake news and "1" for true news. After training the models with the target variable, the output of the test data will be displayed as either 0 or 1 based on the prediction done by the models. In addition to labelling, to achieve better accuracy in fake news detection, columns named Title and Text are grouped and labelled as "news" for both True news and Fake news.

Figure 2: Pre-processing the dataset

4.3 Feature selection

The next step is to select the features that will be used to train the machine learning model. These features can include things like the length of the article, the number of words, the use of certain words or phrases, and the sentiment of the text.

4.4 Model Architecture

TF-IDF and count vectors were the main focus of the study to extract significant features from the Fake News Dataset. A structured pipeline was used to apply an Support Vector Machine model.

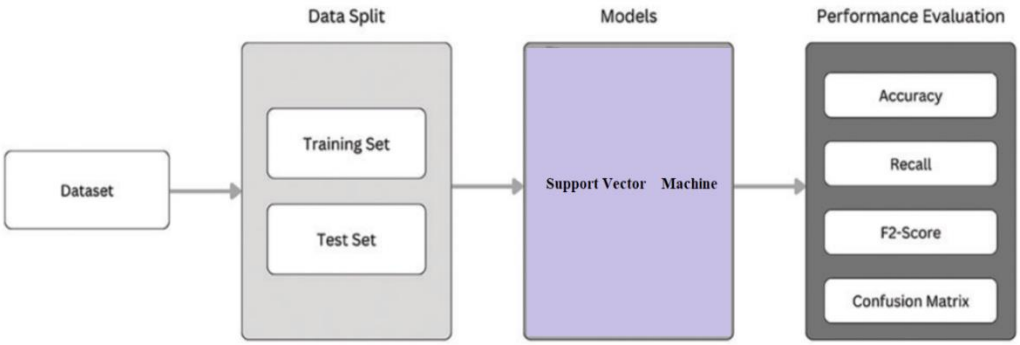

Figure 1: Implementation and Evaluation Process

4.5 Model Training

After the features have been extracted, a model can be trained on the data. There are many different types of models that can be used for fake news analysis, including decision trees, support vector machines, and neural networks. The model is trained on a subset of the data and validated on another subset to ensure that it is accurately identifying fake news

4.6 Model Evaluation

Once the model has been trained, it needs to be evaluated to determine how well it is performing. This can be done using metrics such as precision, recall, and F1 score. These metrics measure the model's ability to correctly identify fake news and avoid false positives.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative}}$$

$$\text{Precision} = \frac{True\ Positive}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{True\ Positive}{\text{True Positive} + \text{False Negative}}$$

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

Receiver Operator Characteristic Curve abbreviated as ROC curve is a graphical plot used in binary classifiers to depict its diagnostic ability. ROC curve is formed by plotting the True Positive Rate on the y-axis and the False Positive Rate on the x-axis at different threshold settings.

5. RESULTS AND FINDINGS

After converting the data into vector form using TF-IDF vector, the model built using Support Vector Machine Classifier algorithm is trained with 80% of data to differentiate Fake news from True News and the remaining 20% of the data are used to evaluate the performance of the model. Model Performance is measured in terms of Confusion Matrix, Classification report and ROC curve. Figure 3 shows the confusion matrix and classification report obtained for the Support Vector Machine Classifier. The confusion matrix clearly illustrates that the model correctly predicted 4542 news as fake news and incorrectly predicted 20 true news as fake news.

similarly, the model correctly predicted 4264 news as true news and incorrectly predicted 26 fake news as true news. The classification report explains the overall success rate of the model measured in terms of Accuracy, Precision, Recall and F1-score. The overall accuracy attained for this model is 99.480. The precision score for the fake news and true news is 100% and 99% respectively. The recall score for the fake news and true news is 99% and 100% respectively. F1-score is 99% for both fake news and true news. The diagnostic ability of the model is plotted in the form of a ROC curve are shown in Figure 4. In the plot, the AUC value is equal to 1.00 which reveals that the model is performing well in distinguishing Fake News and True News.

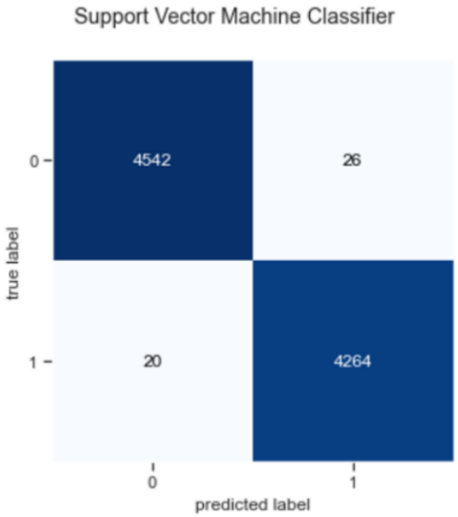

Figure 3: Confusion Matrix of Support Vector Machine Classifier

Table 2: Classification Report of Support Vector Machine Classifier

	Precision	**Recall**	**f1-score**	**Support**
0	1.00	0.99	0.99	4568
1	0.99	1.00	0.99	4284
accuracy			0.99	8852
macro avg	0.99	0.99	0.99	8852
weighted avg	0.99	0.99	0.99	8852

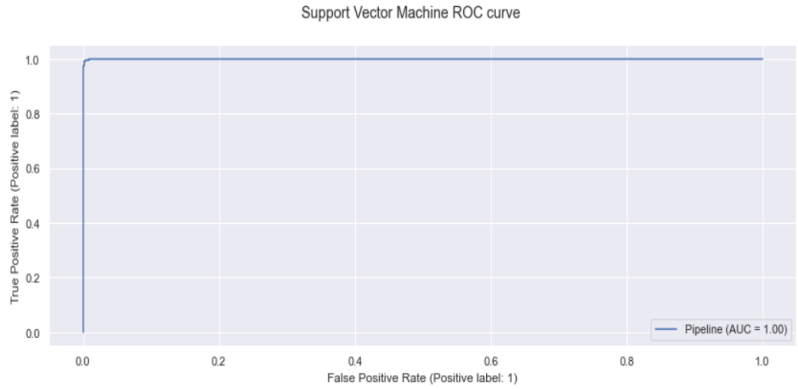

Figure 4: Support Vector Machine ROC Curve

5. DISCUSSION

The results from applying Support Vector Machine to the study of fake news detection are promising, highlighting the effectiveness of this method in distinguishing between false and true news. With an overall accuracy of 99%, the Support Vector Machine model demonstrates a high degree of reliability. The confusion matrix shows a strong performance, with 4542 fake news items correctly identified and only 26 true news items misclassified as fake. Conversely, 4264 true news items were accurately identified, and 20 fake news items were incorrectly classified as true.

6. CONCLUSION

In conclusion, this study successfully employed a Support Vector Machine (SVM) model to classify real or fake with remarkable accuracy. The SVM, integrated with TF-IDF and Count Vectorization for feature extraction, achieved an impressive overall accuracy of 99% on the validation dataset and 99% on unseen test data. Overall,

this SVM model provides a promising solution for mitigating misinformation and ensuring the credibility of information in the digital age.

ACKNOWLEDGEMENTS
Funding Details
This research received no external funding.

Authors' contributions
All authors contributed toward data analysis, drafting and revising the paper and agreed to be responsible for all the aspects of this work.

Declaration of Conflicts of Interests
Authors declare that they have no conflict of interest.

Availability of data and materials
The datasets analyzed during the current study are publicly available at Kaggle.

Declarations
Authors declare that all works are original and this manuscript has not been published in any other journal.

REFERENCE

Abdulqader, D. M., Abdulazeez, A. M. & Zeebaree, D. Q. (2020). Machine Learning Supervised Algorithms of Gene Selection: A Review. Machine Learning 62.03.

Alkhatib, K. et al. (2013). Stock price prediction using k-nearest neighbor (kNN) algorithm. International Journal of Business, Humanities and Technology 3.3: 32-44.

Bharadwaj, P. & Shao, Z. (2019). Fake news detection with semantic features and text mining. International Journal on Natural Language Computing, 8.

Breiman, L., Friedman, J. , Olshen, R. & Stone, C. (1984). Classification and Regression Trees, Springer, Berlin, Germany.

Busioc, C.; Ruseti, S. & Dascalu, M. (2020). A Literature Review of NLP Approaches to Fake News Detection and Their Applicability to Romanian- Language News Analysis, Romanian Ministry of Education and Research, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2019-1794, within PNCDI III.

Cristianini, N. & Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge University Press, Cambridge, UK.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. doi:10.1007/BF00994018

Felber, T. Constraint (2021). Machine Learning Models for COVID-19 Fake News Detection Shared Task. arXiv 2021, arXiv:2101.03717.

Gregorutti, B., Michel, B. & Saint-Pierre, P. (2017). Correlation and variable importance in random forests. Statistics and Computing, vol. 27, no. 3, pp. 659–678.

Hofmann, T., Sch¨olkopf, B. & Smola, A. J. (2008). Kernel methods in machine learning," 7e Annals of Statistics, vol. 36, no. 3, pp. 1171–1220.

Jiang T, Li JP, Haq AU, Saboor A, Ali A (2021) A novel stacking approach for accurate detection of fake news. IEEE Access 9:22626–22639. https:// doi. org/ 10. 1109/ ACCESS. 2021. 30560 79

Kaliyar, R. K. (2018). Fake news detection using a deep neural network. 2018 4th International Conference on Computing Communication and Automation (ICCCA). IEEE

Lazer, D. M. J., Baum, M. A., Benkler, Y., et al. (2018). The science of fake news. *Science*, 359(6380), 1094-1096. doi:10.1126/science.aao2998

Pennycook, G., & Rand, D. G. (2018). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, 115(48), 12401-12406. doi:10.1073/pnas.1811508115

Sawinder, K.; Kumar, P. & Kumaraguru, P. (2020). Automating fake news detection system using multi-level voting model. Soft Computing 24.12, 9049-9069.

Silva, R.M.; Santos, R.L.S.; Almeida, T.A.; Pardo, T.A.S. (2020). Towards Automatically Filtering Fake News in Portuguese. Expert Syst., 146, 113199

Yuan, L.; Jiang, H.; Shen, H.; Shi, L.; Cheng, N. Sustainable Development of Information Dissemination: A Review of Current Fake News Detection Research and Practice. Systems 2023, 11, 458. https://doi.org/10.3390/systems11090458

Zubiaga, A., Liakata, M., & Voss, A. (2016). Learning to detect rumors in social media. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 1751-1756.