International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET229225

# A Survey on Fake News Detection using Support Vector Model

**Himani Kanojia*, Prof. Rahul Vaza**
Computer Engineering Department, Government Engineering College Modasa, Gujarat, India

**Article Info**
Volume 9, Issue 2
Page Number : 125-131

**Publication Issue :**
March-April-2022

**Article History**
Accepted : 20 March 2022
Published: 30 March 2022

**ABSTRACT**

In the boosting period of Social Media availability and easy availability of Internet to end users in various regions, many challenges are also occurred with usage of this technology. Fake news spreading in various field is also a major challenge in recent time. Fake news has been spreading into vast in significant numbers for various business reasons and also for political reasons. Problem of fake news has become frequent in the online world. People can get affected and their view are also affected easily by these type of fake news for its fabricated words. This type of news has enormous effects on the offline community in various sectors. In this way it is an interesting topic for research. Significant research has been conducted on the detection of fake news from English texts and other languages but there is chancel to improve the work with other languages as well as from multiple sources. Various algorithms like SVM and other supervised algorithm can be helpful to classify fake news. As Sentiment Score is an also a major point for detection of Fake news; in our work we are applying SVM algorithm with TF/IDF, multiple Languages (Language Conversation) etc."

Keywords : Artificial Intelligence, Machine Learning, SVM, Text Classification, Sentiment Analysis

## I. INTRODUCTION

News Authentication or authenticity of news is a new challenge in current era of technology. With the ease of technology and social media; people have opted various platforms to get more information. However, an unprecedented amount of the data flooded on the Internet is fake news, which is generated to attract the audience, to influence beliefs and decisions of people to increase the revenue generated by clicking [1], and to affect major events such as political elections [6]. In some cases unethical marketing tactics also use this tool for their marketing. Readers are misguided by deliberately spreading false information. Obtaining and spreading information through social media platforms has become extremely trouble-free, which makes it difficult and nontrivial to detect based merely on the content of news.

To overcome this issue manual checking of each and every news published online is not only difficult but also time consuming as well as spreading of such news is faster, so that is it not advisable for manual check. So that we can say automatic detection of fake news is the need of new era. Natural Language Processing and machine learning technique can work toward this.

In this survey paper we have reviewed some techniques related to fake news detection and also reviewed some research work for the same

## *Various Methodologies*

*Artificial Intelligence:* Artificial Intelligence or sometimes also called machine intelligence, is intelligence demonstrated by machines with data, in contrast to the known intelligence presented by humans or other animals in this world. Some of the activities for that it is designed to do is speech recognition (Text from the speech), learning or guidance, planning for the particular issue and problem solving of the issues. As we know robotics is the popular field in current time period that is concerned with the connection of perception to action. Artificial Intelligence (AI) plays a main role in Robotics. In the field of Science, Artificial Intelligence is used for the crucial problems like what type of knowledge is required in any aspect of thinking. Also how that type of knowledge should be represented if required and how the said knowledge be used. Robotics in the field of Artificial Intelligence is a challenged by forcing it to work with real objects in the real world.

There are various branches in day to day like for AI that plays important role in today's technology. In figure 1 we have presented some of the branches where AI is being used and proving its efficiency. It includes Machine Learning, VISON Technology, NLP (Natural Language Processing), also in Speech and voice sector etc.

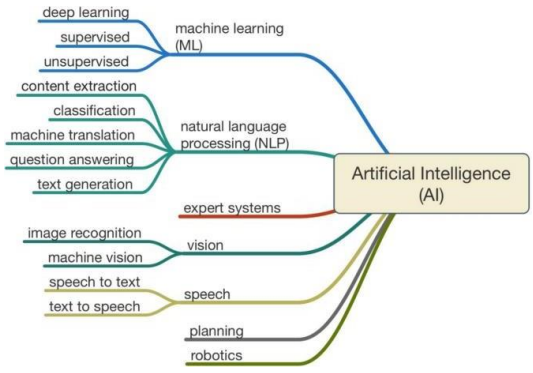


Figure 1 Artificial Intelligence Applications

## *Sentiment Score:*

Sentiment Score is an algorithm result (based on various text of image) for a document or for a Sentence. It can be different for various in figures based on algorithm and techniques of calculation. It is also depends upon for which purpose sentiment score is being calculated and what type of method is applied for keyword and stop words.

The creators or speeders of fake news uses various stylistic tricks to promote the success of their fake news. One of them is to excite the sentiments of the recipients. This point has led to sentiment analysis in fake news detection, the part of text analytics in charge of determining the polarity and strength of sentiments presented in a text of news, to be used in fake news detection approaches, either as a basis of the system or as a complementary element. So it knowledge based approach for Fake News Detection.

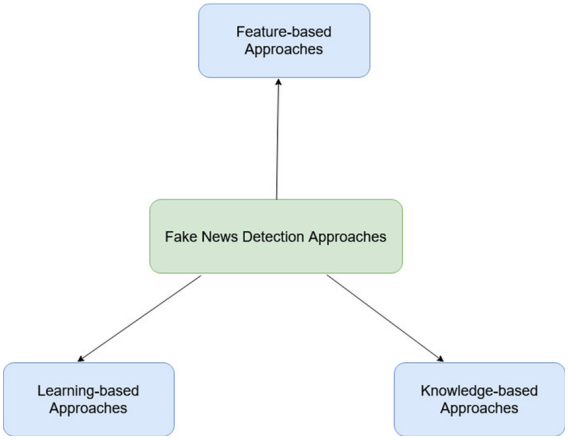


Figure 2 Fake News Detection Approaches

## *SVM*

Support Vector Machine as well-known supervised machine learning algorithm that is used for Classification as well as regression purpose also. There are so many variations of SVM like Weighted SVM, CSVM etc.

The SVM algorithm of Machine Learning calculate a hyper plane decision boundary that is used for best splitting of the given data into two classes. The split in this hyper plane is made soft through the use of a margin. The margin allows some points to be misclassified. By default, this margin favors the majority class on imbalanced datasets, although it can

be updated to take the importance of each class into account and dramatically improve the performance of the algorithm on datasets with skewed class distributions. To overcome this issue Weighted SVM or WSVM can be applied using some weights to the points for classification.

### *Trust Model*

Trust Model [16] is a model that is used by researchers in their work [16] for tweet trust measuring. For this they have applied a system that apply authorship details and also check their historical tweets. Mostly to generate Trust Model authenticity of the published is measures. Generally it is measured based on popularity and history of the social media account. Earlier experience of posts by the account is considered here. Drawback for the Trust model is: if earlier record of the publisher is not available then it can not work.

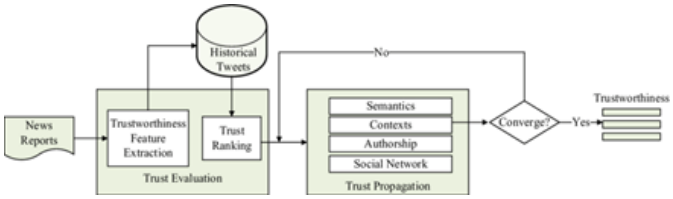


Figure 3 Trust Model [16]

## II. LITERATURE STUDY

We have reviewed some related works related to Fake News Detection and Sentimental Analysis in various region of the world. We have reviewed a research work[1]. In their work, they used two supervised machine learning algorithms, among them one is Multinomial Naive Bayes (MNB) and another algorithm they used is Support Vector Machine (SVM). These classifiers are used detect Bangladesh fake news. They have used CountVectorizer and Term Frequency - Inverse Document Frequency Vectorizer as Feature extraction. Their proposed framework detects fake news depending on the polarity of the corresponding new contents. Finally, their analysis proved that SVM algorithm with the linear kernel with an accuracy of 96.64% outperform MNB with an accuracy of 93.32%.

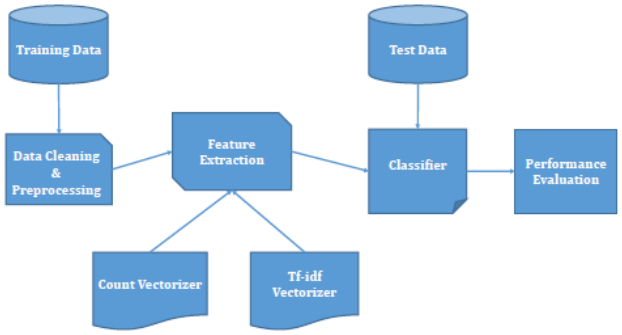


Figure 4 System Flow for Fake News Detection[1]

In research work "CNN-LSTM[2]" researchers have used CNN and LSTM as a part of their research. It is used with two different dimensions and reduction approaches. These two dimensions are Principle Component Analysis (PCA) and Chi-Square. Their work was proposed to employ the dimensionality reduction techniques. It aims to reduce the dimensionality of the feature vectors. It is reduced before passing them to the classifier. For their research work, this work acquired a dataset from the Fake News Challenges (FNC) website. Their dataset has four types of stances. The stances contains agree, disagree, discuss, and unrelated. The nonlinear features in the work are fed to PCA and chi-square. These dimension provides more contextual features for fake news detection.

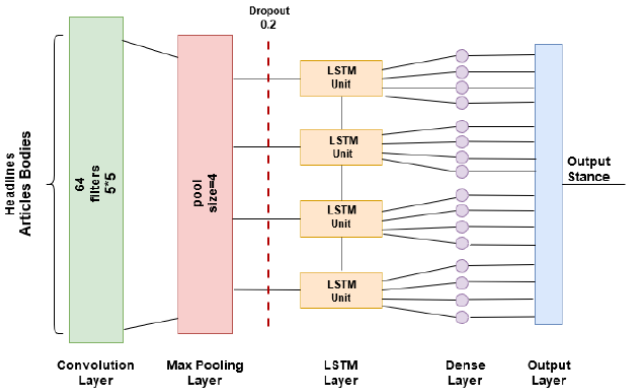


Figure 5 Architecture of CNN-LSTM Model[2]

In work [3], researchers have designed an effective deep neural network for Fake News Detection. The framework is capable of handling not only the content of the news but also the relationships of the users in the social network. They have designed the proposed approach of the framework using tensor factorization method. In general a tensor expresses the social context of news articles. This is formed by a combination of the news, user, and user-group information. They have applied Fakeddit Dataset. This dataset is derived from Fake News + Sources of Reddit. It contains 8,00,000 examples of fake news.

In research work[4] NLP Techniques are applied for fake news detection in the work. They achieved 83% AU-ROC. They have used Korean Model for training. In research work[5] simple approach using KNN is applied. They have used Facebook Social Media post as their testing dataset and got 79% Accuracy using KNN Model.

In research work[6], researchers have considered Sentiments of the News as a strong feature to detect the news fake or real. As per their opinion Sentiments of the text can be a decisive feature. Also they applied TF-IDF Vectorization.

After reviewing these research work we have summarized the work in following table

| Research Paper | Findings |
|---|---|
| [1] | Their research work shows the experimental analysis on the detection of fake news from Bangladesh in various social media.<br>They have used 2 supervised ML algorithms.<br>One is Multinomial Naive Bayes (MNB) and second is Support Vector Machine (SVM) classifiers.<br>Accuracy of 96.64% achieved using SVM and MNB with an accuracy of 93.32% |
| [2] | In this research a hybrid NN architecture is applied.<br>Combination of CNN and LSTM is used.<br><br>Fake News Challenges (FNC) website dataset is used.<br>The proposed model improves results by 4% in terms of Accuracy |
| [3] | Tensor factorization method is used.<br>Combination of the news, user, and user-group information is used.<br>Real-world fake news dataset: BuzzFeed and Fakeddit are tested. |
| [4] | They performed various NLP tasks, and create a Korean specific pre-training model using state of the art BERT algorithm.<br>The AUROC score as a result of 83% is achieved from the test set generated using the fine-tuned model |
| [5] | They applied simple approach for detecting fake news on social media with the help of K-Nearest Neighbor (KNN) classifier<br>Results achieved a classification accuracy of this model approximate 79% tested against Facebook news posts dataset |
| [6] | They applied sentiment as an important feature to improve the accuracy<br>Three different data sets are tested.<br>tf-idf Vectorizer is applied. |

| | |
|---|---|
| [7] | They applied different learning experiments from a multilingual perspective, English and Spanish. Different textual features that are primarily not tied to a specific language and compare different machine learning algorithms are applied. |
| [8] | Taxonomy for entities classification is used. Social Media Data are tested. A semantic model is developed to describe classes |
| [9] | Domain reputation analysis is done. Tf-idf & LDA is applied. |
| [10] | Graph-based semi-supervised fake news detection method is applied. It is based on graph neural networks is proposed. Fast Belief Propagation (FaBP) algorithm is applied. |

## III. COMPARATIVE STUDY

This section will go through the most commonly used machine learning algorithms and transfer learning algorithms, as well as their advantages. There are other disadvantages.

| Method | Advantages | Disadvantages |
|---|---|---|
| Linear Regression [2,3,14] | Easy and simple implementation. Space complex solution. | Applicable only if the solution is linear. In many real life scenarios, it may not be the case. Algorithm assumes the input residuals (error) to be normal distributed, but may not be satisfied always. |
| SVM[1,4,7,11] | When data has high dimension a Support Vector Machine with the right settings. Support Vector Machines may be relatively sluggish when it comes to training especially with large datasets, for prediction they are quite faster. | SVMs are not the most efficient algorithms and it can be quite costly computationally to train them. Dataset in hand will already need have feature vectors or you will need to pre-process to extract feature vectors which might not always be easy or possible. |
| Naïve Bayes[1,7] | This algorithm works very fast and can easily predict the class of a test dataset. Naive Bayes classifier performs better than other models with | If your test data set has a categorical variable of a category that wasn't present in the training data set, the Naive Bayes model will assign it zero probability and won't be able to make any predictions in this regard. |

| | | |
|---|---|---|
| | less training data if the assumption of independence of features holds. | |

## IV. CONCLUSION AND FUTURE DIRECTION

After reviewing so many research work related to fake news detection, we have concluded that the issue of detection of Fake News can be handled with better classification using Machine Learning Algorithms and NLP. Multiple Source verification can help for better result.

As a part of future work there is chance to improve the overall result by applying multiple resource authentication or extra features of news. Some trust based authentication may also be helpful to identify the fake and real news in proper way.

## V. REFERENCES


[1]. Hussain, Md Gulzar, et al. "Detection of bangla fake news using mnb and svm classifier." 2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE). IEEE, 2020.

[2]. M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi and B. -W. On, "Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM)," in IEEE Access, vol. 8, pp. 156695-156706, 2020, doi: 10.1109/ACCESS.2020.3019735.

[3]. R. K. Kaliyar, P. Kumar, M. Kumar, M. Narkhede, S. Namboodiri and S. Mishra, "DeepNet: An Efficient Neural Network for Fake News Detection using News- User Engagements," 2020 5th International Conference on Computing, Communication and Security (ICCCS), 2020, pp. 1-6, doi: 10.1109/ICCCS49678.2020.9277353.

[4]. Y. Ahn and C. Jeong, "Natural Language Contents Evaluation System for Detecting Fake News using Deep Learning," 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2019, pp. 289-292, doi: 10.1109/JCSSE.2019.8864171.

[5]. A. Kesarwani, S. S. Chauhan and A. R. Nair, "Fake News Detection on Social Media using K-Nearest Neighbor Classifier," 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), 2020, pp. 1-4, doi: 10.1109/ICACCE49060.2020.9154997.

[6]. B. Bhutani, N. Rastogi, P. Sehgal and A. Purwar, "Fake News Detection Using Sentiment Analysis," 2019 Twelfth International Conference on Contemporary Computing (IC3), 2019, pp. 1-5, doi: 10.1109/IC3.2019.8844880.

[7]. I. Vogel and M. Meghana, "Detecting Fake News Spreaders on Twitter from a Multilingual Perspective," 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 2020, pp. 599- 606, doi: 10.1109/DSAA49011.2020.00084.

[8]. Bani-Hani, Anoud I. et al. “A Semantic Model for Context-Based Fake News Detection on Social Media.” 2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA) (2020): 1-7.

[9]. K. Xu, F. Wang, H. Wang and B. Yang, "Detecting fake news over online social media via domain reputations and content understanding," in Tsinghua Science and Technology, vol. 25, no. 1, pp. 20-27, Feb. 2020, doi: 10.26599/TST.2018.9010139.

[10]. A. Benamira, B. Devillers, E. Lesot, A. K. Ray, M. Saadi and F. D. Malliaros, "Semi-Supervised Learning and Graph Neural Networks for Fake News Detection," 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2019, pp. 568-569, doi: 10.1145/3341161.3342958.

[11].https://towardsdatascience.com/using-a-i-to-combat- fake-news-34f5a51907d6
[12].Dataset, https://www.kaggle.com/c/fake-news/data
[13].https://www.researchgate.net/profile/Ziyad_Saleh2/p ublication/332548325_Artificial_ Intelligence_ Definition_Ethics_and_Standards/links/5cbc3cc492851 c8d22fad1a6/Artificial-Intelligence-Definition-Ethics- and-Standards
[14].Kaliyar, Rohit & Goswami, Anurag & Narang, Pratik. (2021). DeepFakE: improving fake news detection using tensor decomposition-based deep neural network. The Journal of Supercomputing. 77. 10.1007/s11227-020-03294-y.
[15].https://miro.medium.com/max/602/0*qf-O7Jm1mmZrXYqA
[16].Zhao, Liang & Hua, Ting & Lu, Chang-Tien & Chen, Ing- Ray. (2015). A Topic-focused Trust Model for Twitter. Computer Communications. 14. 38-1. 10.1016/j.comcom.2015.08.001.

**Cite this article as :**

Himani Kanojia, Prof. Rahul Vaza, "A Survey on Fake News Detection using Support Vector Model", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 2, pp. 125-131, March-April 2022. Available at doi : https://doi.org/10.32628/IJSRSET229225
Journal URL : https://ijsrset.com/IJSRSET229225