



## A MACHINE LEARNING BASED APPROACH FOR IMPROVED FAKE NEWS DETECTION

Atul Suryawanshi

M.Tech Scholar, Department of CSE  
Radharaman Institute of Technology & Science  
Bhopal, India

Vijendra Palash

Asst. Prof., Dept. of CSE  
Radharaman Institute of Technology & Science  
Bhopal, India

Priyank Nayak

Asst. Prof., Dept. of CSE  
Radharaman Institute of Technology & Science  
Bhopal, India

**Abstract:** The News is significant piece of our life. In everyday life current news are useful to improve information what occur all throughout the planet. So the vast majority of people groups lean toward watching news a large portion of the people groups for the most part favor perusing paper promptly toward the beginning of the day appreciating with cup of tea. On the off chance that news is phony that will delude people groups now and then phony word used to get out bits of gossip about things or it will influence some political pioneer positions on account of phony news. So it's vital to track down the phony news. This exploration proposed an advanced framework to distinguish counterfeit news, yet now daily's information on web or online media is expanding immensely and it is so rushed to recognize news is phony or not by looking all information and it is tedious so we use characterization strategies to order colossal information. This paper proposed fake news detection system based on the classification approach such as Naïve bayes (NB), Support vector machine (SVM), K Nearest Neighbor (KNN) and Decision Tree (TD)

**Keywords:** Machine Learning, Fake News, KNN, DT, NB, SVM, Python.

### I . INTRODUCTION

The News is indispensable piece of our life. In regular day to day existence current news are useful to improve information what happen the world over. So an enormous segment of social classes favor watching news most of the social classes generally favor scrutinizing paper expeditiously around the start of the day getting an accuse out of cup of tea. In the event that news is phony that will bamboozle social classes here and there counterfeit word used to get out gossipy goodies about things or it will influence some political pioneer positions by virtue of phony news. So it's urgent to track down the phony news[1]. So we proposed system to recognize counterfeit news however at this point day by day's data on web or web-based media is growing interminably and it is so hot to distinguish news is phony or not by looking all data and it is time consuming so we use grouping strategies to characterize enormous data.

Highlights of society draw in the progress of meaning of those procedures for friendship that is considered the contrast in larger part's insights with respect to fighting toward a way that is profitable for one of the clashing parties. Such changes are conceivable by virtue of extraordinarily picked information spreading first through the far-reaching interchanges and through channels of nice correspondence. Pieces of tattle for this situation brings certified weapon up in conflicts. Issue of perceiving or discovering bogus data in customary everyday presence, albeit particularly like trickiness finding, yet it is basic to distinguish because of the news body often includes a

less and little declarations. Issues related to such subjects are much of the time been seen relying upon order. In like manner, an enormous part of printed matters has considered bogus to be area as a twofold order issue. Some association utilizes various profound learning strategies on data sets made out of bogus data articles and certified data articles mined from media news information base and saw classifiers is incredible to group immense information.

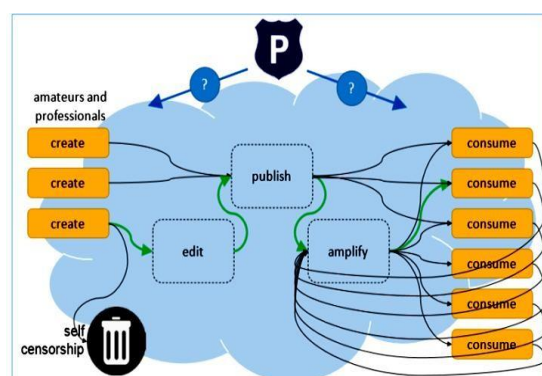


Figure 1: News process in the age of the internet

The path toward getting news from online media resembles twofold edged weapon. On one hand, it is definitely not hard to get to, less time eating up, simple to

utilize, successfully conveyable socially appropriate news, openings for securing diverse perspective of singular news and is being invigorated in reliably. On other hand, news is being constrained by various systems administration objections subject to private notions or interest. Counterfeit news is deception or controlled word that is gotten out over the web-based media with an intend to hurt an individual, office and affiliation. As a result of the dispersal of phony news, there is necessity for computational procedures to remember them. Counterfeit news acknowledgment intends to help customers with revealing groupings of manufactured news.

In This paper contains are literature Review given in 2, in 3 Proposed Methodology is discussed, Simulation Results are given in 4, and lastly, the conclusion is to describe.

## II. LITERATURE REVIEW

**N. Smitha et al.,[2]** Internet information exchanges, especially online media information exchanges, increased by person-times. Accidental or deliberately generated false information is generated through the Internet . Society is blinded by technology. This document explains a model and the method of using machine learning to detect fake news in news articles. different Machine learning Classification algorithms are trained to classify news as fake or real and are compared considering accuracy.

**S. I. Manzoor et al.,[3]** the research task began to automatically check the information, namely the source, content, and publisher, classifying it as false or true. Machine learning has played a vital role in information classification and has some limitations. The article examines different machine learning methods to identify false and fabricated news.

**S. Lyu et al.,[4]** Research on detecting fake news has been around for several years. It is a relatively new and complex research field, which is difficult due to the semantics of natural language and manual recognition by humans, let alone machines. The performance of various machine learning algorithms combined with support vector machines (SVM) and decision trees. Our preliminary results show that SVM and decision trees can detect fake news with an acceptable accuracy rate of 95%. Usually, decision tree methods do better than SVM. Future research directions are taken into consideration.

**W. Antoun et al.,[5]** This introduces modern methods to solve the three main problems of automatic fake news detection: fake news detection, domain recognition, and robot recognition in tweets. A deep learning model was used to identify style differences between legitimate and fake news articles. The following characteristics: the duration between the account creation and the tweet date, the existence of the tweet link, the existence of the user location, other characteristics of the tweet, and the metadata of the tweet experiment contains the meaning of various attributes of the relevant information, and the results indicate The superiority of the attributes of all the proposed models.

**S. B. Parikh et al.,[6]** Fake news has existed for decades, and as social media and modern news have reached their peak, fake news detection has become a hot topic in the research community. The world is trying to understand the main

characteristics of the problem. This article aims to combine different types of news content and their impact on readers and demonstrate the characteristics of news among the contemporary. Existing fake news detection methods are mainly based on text analysis, and also describe popular fake news data sets.

**M. Kumar Jain et al.,[7]** Most users on social networks do not check and distribute information. Manually identifying fake news is a big problem for everyone. This article evaluates a model that can visually distinguish fake news from news articles. A new set of machine learning classifier functions is proposed. In the experiment, the data set used is a combination of two data sets containing the same real news and political fake articles. Extract language/stylistic features, text fields of the data set, and then apply various machine learning models, including packing and reinforcement methods.

**K. Poddar et al.,[8]** Fake news is news that has not been fully researched or is intentionally used to spread misinformation or deceptive news through various forms of news distribution networks. This article aims to solve this problem using computational modeling, probability machine learning. In order to find suitable victories for detecting fake news, the values of two different victories were compared. English stop words were used to improve the results. Various classifiers have been used to predict fake news, such as B. Naive Bayes analysis, support vector machines (SVM), logistic regression and decision tree classifiers.

**D. K. Sharma et al.,[9]** Disclosure of fake news is a requirement of the modern digital age. In this document, we discuss different ways to discover misleading content that the public may be viewing. We also detailed existing tools and extensions that can be used to detect fake news. We showed some systems designed by researchers to combat false information. Various fact-checking websites are analyzed here to help social media users check the information provided on social media. This document aims to help the public learn the basic methods of detecting fake news. the existing Kaggle data set and used the BiLSTM classifier to achieve an accuracy of 91.51%.

**I. Vogel et al.,[10]** Fake news creators often use false information from trusted news sources to intentionally or unknowingly mislead readers. This is increasingly seen as a threat to democracy, public order and free debate, and can lead to chaos and unrest. Real-time feedback is very important. Usually, it is not enough to quickly check the content of all sent messages. Spreading fake news on social media is the first step to prevent fake news from spreading to online users. Language-independent features can be used to distinguish potential fake news from users who share valid information. , The average recognition accuracy rate is 78%.

**V. V. Hirlekar et al.,[11]** Internet social media plays an important role in real-life events such as natural disasters, elections, and social movements the production and dissemination of fake news pose a serious threat in several aspects. Therefore, the identification of fake news is becoming an important goal to improve the reliability of information disseminated on social media. Many researchers use various

methods, algorithms, tools, and techniques to identify fake news content on social media. This plays an important role in achieving maximum accuracy through various machine learning and natural language processing algorithms.

### III. PROPOSED METHODOLOGY

**Steps:-**

- Firstly, download the dataset from kaggle website, which is a largedataset provider company for research.
- Now preprocessing of the data, here handing the missing dataset. Remove the null value or replace from common 1 or 0 value.

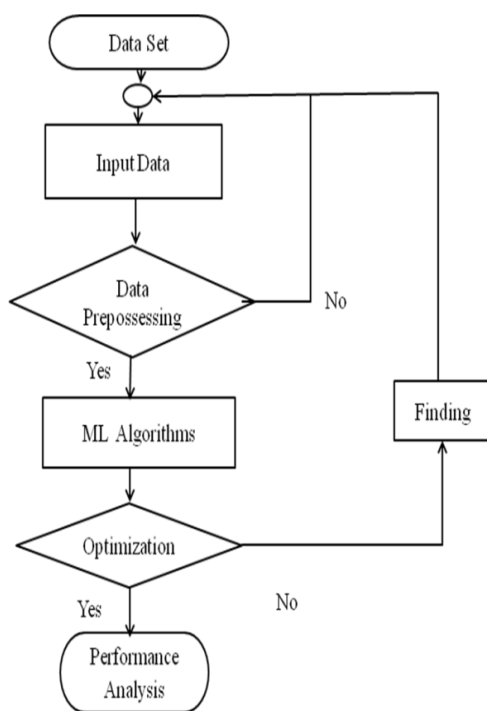


Figure 2: Flow chart of Proposed Method

- Now apply the classification method based on the machine learning approach. The Naïve bayes (NB)[13], Support vector machine (SVM)[12], K Nearest Neighbor (KNN) and Decision Tree (DT) machine learning method is applied[14].
- Now check and calculate the performance parameters in terms of the precision, recall, F\_measure, accuracy and error rate.
- It incorporates expulsion of punctuations, URL's, images, stemming and stop words.
- At that point classify that information utilizing classifiers, for example, NB, SVM, KNN and DT.
- News Dataset: Dataset of news is taken from Kaggle website and size of data is 14000 .
- Processing: Content information needs processing to execute AI on them. There are kinds of methods

generally utilized to change over content Information into a structure that is prepared for demonstrating. The information processing steps that are applied.

### IV. SIMULATION AND RESULT

The implementation of the proposed algorithm is done over python spyder 3.6. The sklearn, numpy, pandas, matplotlib, pyplot, seaborn, os library helps us to use the functions available in spyder environment for various methods like support vector, random forest, naive bayes, DT etc.

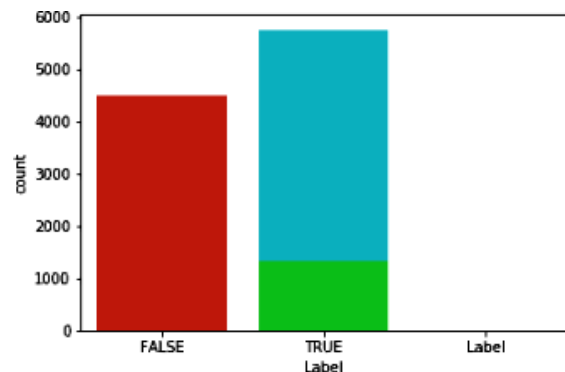


Figure 3: News label

Figure 3 is showing the NEWS label, here are showing false news, true news label. The lier data set is taken from various news.

news_no	user_spread_news_no_times
44	45
31	32
179	180
139	140
152	153
92	93
86	87
167	168
41	42
71	72
170	171
90	91
94	95
5	6
125	126
180	181
137	138
126	127
63	64
149	150
45	1396
32	1228
180	1314
140	1193
153	957
93	840
87	823
168	744
42	735
72	734
171	648
91	627
95	552
6	521
126	440
181	412
138	385
127	379
64	356
150	337

Figure 4: Sample of dataset

Figure 4 is showing the dataset. The total training set is taken 14000.

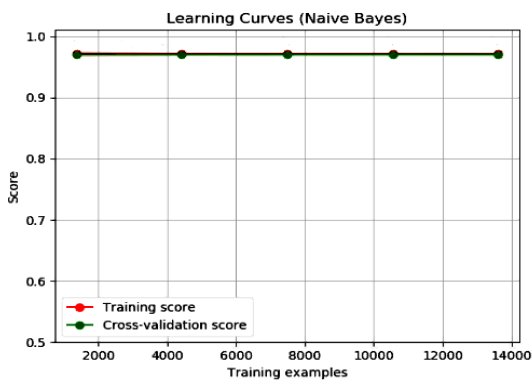


Figure 5: learning curves of naïve bayes (NB)

Figure 5 is showing the learning curve of naïve bayes. Results show that the naïve bayes algorithm gives 95% accuracy during training score and cross validation score.

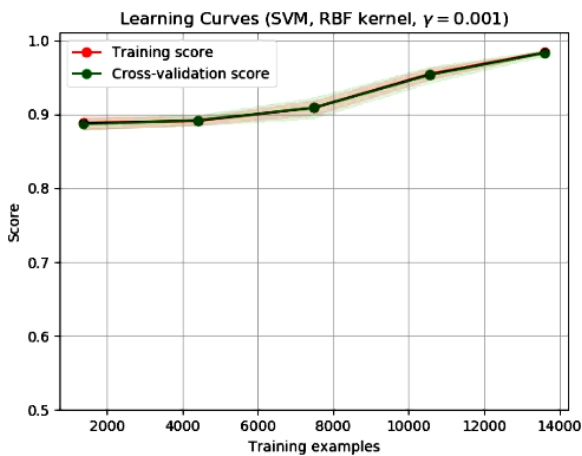


Figure 6: learning curves of Support Vector Machine (SVM)

Figure 6 is showing the learning curve of support vector machine. Results show that the Support Vector Machine gives 88% accuracy during training score and cross validation score.

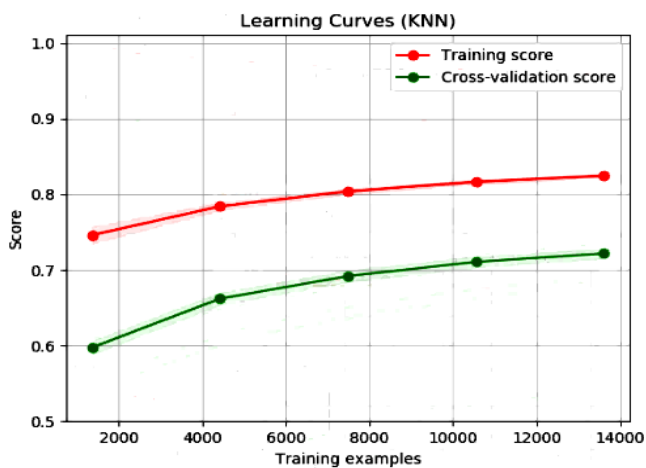


Figure 7: learning curves of K Nearest Neighbor (KNN)

Figure 7 is showing the learning curve of K Nearest Neighbor. Results show that the K Nearest Neighbor algorithm gives 82% accuracy during training score and cross validation score.

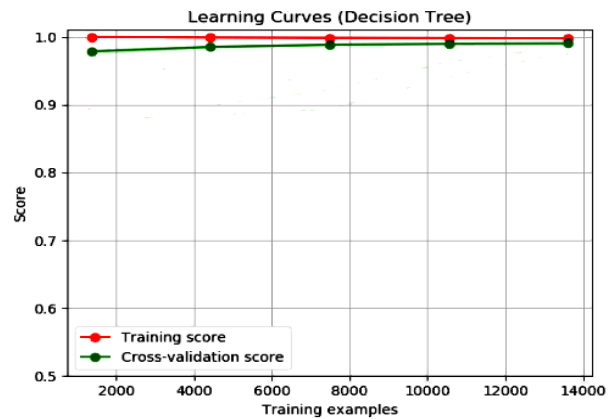


Figure 8: learning curves of decision tree (DT)

Figure 8 is showing the learning curve of decision tree algorithm. The total training set is taken 14000. Results show that the decision tree algorithm gives maximum accuracy i.e. 98% during training score and cross validation score.

Table 1: Simulation Results

Sr.No.	Methodology	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)	Time (Sec)
1	Naïve Bayes (NB)	96	94	96	95	0.79
2	Support Vector Machine (SVM)	88	89	87	88	0.3
3	K Nearest Neighbor (KNN)	82	83	81	82	2
4	Decision Tree (DT)	98	96	97	98	0.32

Table 2: Comparison of previous work and proposed work

Sr No	Parameters	Previous Work	Proposed Work
1	Method	Deep Learning	Decision Tree
2	Accuracy	91%	98%
3	Error Rate	9%	2%

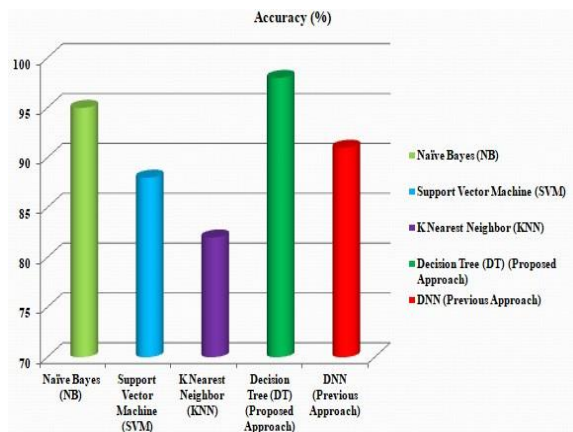


Figure 9: Comparison of accuracy

Figure 9 is showing the comparison of accuracy of various methods. The decision tree method achieves the maximum accuracy that is 98%.

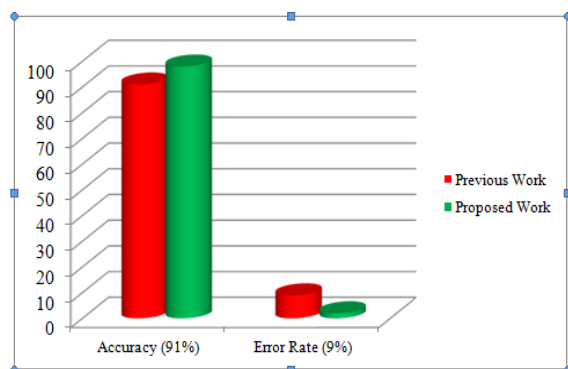


Figure 10: Comparison of previous and proposed accuracy and error rate

Figure 10 is showing the graphical bar chart comparison of previous and proposed method in terms of accuracy and error rate.

## V. CONCLUSION

This paper presents different algorithms for classifying statements made by public figures were implemented. In proposed system Naïve bayes (NB), Support vector machine (SVM), K Nearest Neighbor (KNN) and Decision Tree (DT). We compare all machine learning techniques for detecting fake news. Simulation is performed using Python Spyder 3.6 software. Results shows that proposed decision tree method achieves the maximum accuracy that is 98%.

In the future work, by using the word ontology with the help of deep learning we can improve the accuracy of fake news detection can be done.

## VI. REFERENCESE

- [1] S. Helmstetter and H. Paulheim, "Weakly Supervised Learning for Fake News Detection on Twitter," 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2018, pp. 274-277, doi: 10.1109/ASONAM.2018.8508520.
- [2] N. Smitha and R. Bharath, "Performance Comparison of Machine Learning Classifiers for Fake News Detection," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 696-700, doi: 10.1109/ICIRCA48905.2020.9183072.
- [3] S. I. Manzoor, J. Singla and Nikita, "Fake News Detection Using Machine Learning approaches: A systematic Review," 2019 3rd International Conference on Trends in Electronics and Informatics(ICOEI),2019,pp.230-234,doi: 10.1109/ICOEI.2019.8862770.
- [4] S. Lyu and D. C. -T. Lo, "Fake News Detection by Decision Tree," 2020 SoutheastCon, 2020, pp. 1-2, doi: 10.1109/Southeast Con 44009.2020.9249688.
- [5] W. Antoun, F. Baly, R. Achour, A. Hussein and H. Hajj, "State of the Art Models for Fake News Detection Tasks," 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIOT), 2020, pp. 519-524, doi: 10.1109/ICIOT48696.2020.9089487.
- [6] S. B. Parikh and P. K. Atrey, "Media-Rich Fake News Detection: A Survey," 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2018, pp. 436-441, doi: 10.1109/MIPR.2018.00093.
- [7] M. Kumar Jain, D. Gopalani, Y. Kumar Meena and R. Kumar, "Machine Learning based Fake News Detection using linguistic features and word vector features," 2020 IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2020, pp. 1-6, doi: 10.1109/UPCON50219.2020.9376576.
- [8] K. Poddar, G. B. Amali D. and K. S. Umadevi, "Comparison of Various Machine Learning Models for Accurate Detection of Fake News," 2019 Innovations in Power and Advanced Computing Technologies (i-PACT), 2019, pp. 1-5, doi: 10.1109/i-PACT44901.2019.8960044.
- [9] D. K. Sharma, S. Garg and P. Shrivastava, "Evaluation of Tools and Extension for Fake News Detection," 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM), 2021, pp. 227-232, doi: 10.1109/ICIPTM52218.2021.9388356.
- [10] 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.
- [11] V. V. Hirlekar and A. Kumar, "Natural Language Processing based Online Fake News Detection Challenges – A Detailed Review," 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 748-754, doi: 10.1109/ICCES48766.2020.9137915.
- [12] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," 2019 Fourth International Conference on Informatics and Computing (ICIC), 2019, pp. 1-5, doi: 10.1109/ICIC47613.2019.8985884.
- [13] W. Le, M. Lin, L. Jia, J. Ai, X. Fu and Z. Chen, "Multi-Objective Optimization of an Air-Cored Axial Flux Permanent Magnet

Synchronous Machine with Segmented PMs based on Support Vector Machine and Genetic Algorithm," 2019 22nd International Conference on Electrical Machines and Systems (ICEMS), 2019, pp. 1-4, doi: 10.1109/ICEMS.2019.8922465.

[14] M. Murugappan et al., "Facial Expression Classification using KNN and Decision Tree Classifiers," 2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP),2020,pp.1-6, doi: 10.1109/ICCCSP49186.2020.9315234.