



**IJIRCCCE**

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

**Volume 9, Issue 7, July 2021**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.542**



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

# Fake News Detection Using Machine Learning Algorithm

**Dr.Sasikala, N.Ezhilarasi, A.Farhath Saheera, R.Rathna Priyanga**

Associate Professor, Department of Computer Science and Engineering, Velammal College of Engineering and Technology, Madurai, Tamil Nadu, India

UG Students, Department of Computer Science and Engineering, Velammal College of Engineering and Technology, Madurai, Tamil Nadu, India

**ABSTRACT:** Most of the smart phone users prefer to read the news via social media over internet. The news websites are publishing the news, providing the source of authentication. Human inefficiency to distinguish between true and false facts poses fake news as a threat to logical truth, which deteriorates democracy, journalism, and credibility in governmental institutions. In the wake of emerging technologies, there is dire need to develop methodologies, which can minimize the spread of fake messages or rumours that can harm society in any manner. It is harmful for the society to believe on the rumours and pretend to be a news. The need of an hour is to stop the rumours especially in the developing countries, and focus on the correct, authenticated news articles.

## I. INTRODUCTION

The Internet has become the most powerful tool for communication. It facilitates efficient and effective transfer of media from one location to another. Lie gets travelled around us quicker, and more extensively than reality in all spheres of information, and the effects were more dangerous and horrifying. As fast the technology is moving, on the same pace the preventive measures are required to deal with such activities. The fight against fake news renders the social network and data consumption problems inseparable. By spreading malicious content, a user is wasting network and processing resources and undermining the credibility of the service provided. Fake news on social media which got viral like a rocket in no time can cause much havoc to our society human and country.

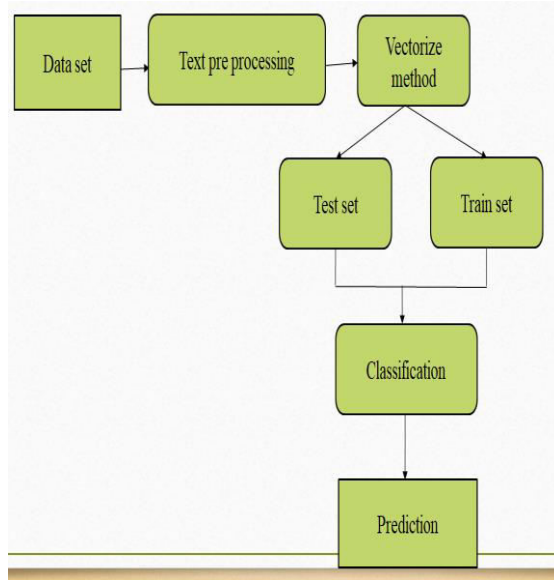
## II. OBJECTIVE

1. To effectively classify and predict the data.
2. To decrease sparsity problem.
3. To enhance the performance of the overall prediction results.
4. The main objective is to detect the fake news, which is a text classification problem. It is needed to build a model that can differentiate between "Real" and "Fake" news.

## III. PROPOSED SYSTEM

1. We are applying data mining techniques to identify suitable techniques for predicting the real and fake news.
2. This system will increase the accuracy of the Supervised classification results by classifying the data based on the text content and others using classification algorithm.
3. It enhances the performance of the overall classification results

#### IV. ARCHITECTURE

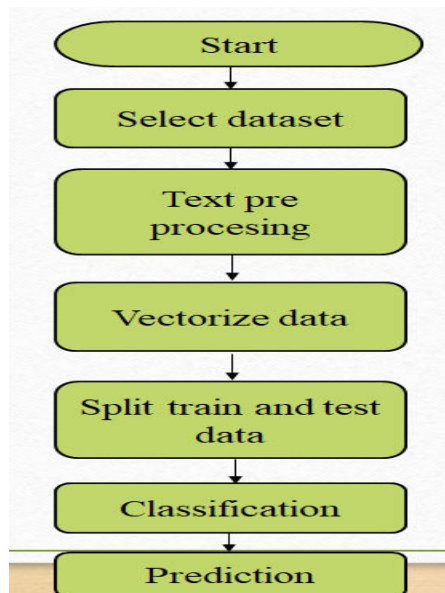


i) architecture diagram of fake news detection

#### V. TECHNOLOGIES USED

1. Python
2. Machine learning

#### VI. FLOW CHART



ii) flow chart diagram of fake news detection

#### VII. IMPLEMENTATION DETAILS AND PERFORMANCE RESULT

##### Data selection and loading:

- The data selection is the process of selecting the data for detecting the news.
- In this project, the news\_articles dataset is used for detecting the real and fake news.
- The dataset which contains the information about the title, text, language, type and label



**Data preprocessing:**

- Data pre processing is the process of removing the unwanted data from the dataset.
  - ❖ Missing data removal
  - ❖ Encoding Categorical data
- Missing data removal: In this process, the null values such as missing values and Nan values are replaced by 0.
- Encoding Categorical data: That categorical data is defined as variables with a finite set of label values. That most machine learning algorithms require numerical input and output variables. That an integer and one hot encoding is used to convert categorical data to integer data.

**Splitting Dataset into Train and Test Data:**

- Data splitting is the act of partitioning available data into two portions, usually for cross-validator purposes.
- One Portion of the data is used to develop a predictive model and the other to evaluate the model's performance.
- Separating data into training and testing sets is an important part of evaluating data mining models.
- Typically, when you separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing.

**Classification:**

- The Supervised classification algorithm such as Naïve Bayes, Random forest and Support vector machine is used in Data Mining.
- **Decision Tree Mining** algorithm is a type of data mining technique that is used to build Classification Models. It builds classification models in the form of a tree-like structure and it can be used for both categorical and numerical data.
- **Gradient boosting** algorithm is a greedy algorithm and can over fit a training dataset quickly. It can benefit from regularization methods that penalize various parts of the algorithm and generally improve the performance of the algorithm by reducing over fitting.

**VIII. RESULTS**

```

47110
(2045, 47110)
PERFORMANCE ANALYSIS
-----
Decision Tree
-----Classification Report-----
              precision    recall  f1-score   support

     0           0.81         0.81         0.81         263
     1           0.66         0.66         0.66         146

   micro avg           0.76         0.76         0.76         409
   macro avg           0.74         0.74         0.74         409
weighted avg           0.76         0.76         0.76         409

-----Accuracy-----
The Accuracy Score :76.0
-----
    
```

iii)results obtained from decision tree classification

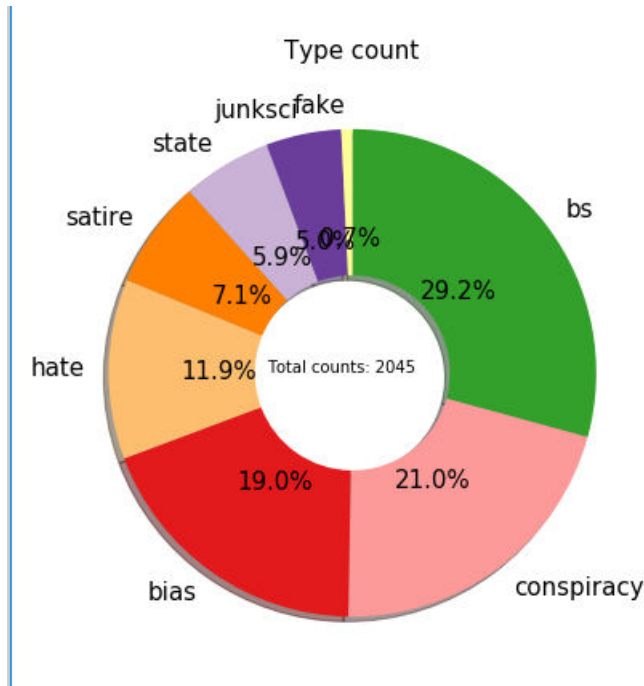


```

-----
Gradient Boosting
-----Classification Report-----
              precision    recall  f1-score   support

|
|         0         0.92      0.80      0.86         304
|         1         0.58      0.80      0.67         105
|
|    micro avg         0.80      0.80      0.80         409
|    macro avg         0.75      0.80      0.76         409
|   weighted avg         0.83      0.80      0.81         409
|
|-----Accuracy-----
| The Accuracy Score :80.0
    
```

iv) results obtained from gradient boosting algorithm



v) pie chart representing the analysis of news from article

### IX. CONCLUSION

- ✓ The model must efficiently distinguish between true news and fake news with maximum accuracy.
- ✓ The existing model's accuracy score will be nearly 80%.
- ✓ The accuracy can be increased with our model.

### REFERENCES

1. Wenlin Han and Varshil Mehta "Fake News Detection in Social Networks Using Machine Learning and Deep Learning: Performance Evaluation" Published in: 2019 IEEE International Conference on Industrial Internet (ICII)
2. S. Sommariva, C. Vamos, A. Mantzarlis, L. U.-L. Dào, and D. Martinez Tyson, "Spreading the (fake) news: exploring health messages on social media and the implications for health professionals using a case study," Amer. J. Health Edu., vol. 49, no. 4, pp. 246–255, Jul. 2018, doi: 10.1080/19325037.2018.1473178.





3. G. Whitehouse, "Pete/Repeat Tweet/Retweet Blog/reblog: A hoax reveals media mimicking," *J. Mass Media Ethics*, vol. 27, no. 1, pp. 57–59, Jan. 2012.
4. M. Kosfeld, "Rumours and markets," *J. Math. Econ.*, vol. 41, no. 6, pp. 646–664, Sep. 2005.
5. Y. Xiao, D. Chen, S. Wei, Q. Li, H. Wang, and M. Xu, "Rumor propagation dynamic model based on evolutionary game and antirumor," *Nonlinear Dyn.*, vol. 95, no. 1, pp. 523–539, Jan. 2019.
6. V. Banerjee, "The economics of rumours," *Rev. Econ. Stud.*, vol. 60, no. 2, pp. 309–327, Apr. 1993.



**INNO**  **SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 7.542**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
**INDIA**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details