

Fake News Detection Using Supervised Machine Learning

Shivansh Singh

School of Computer Science and
Engineering
Galgotias University
Greater Noida, Uttar Pradesh, India
shivansh.singh000@gmail.com

Manish Rajput

School of Computer Science and
Engineering
Galgotias University
Greater Noida, Uttar Pradesh, India
rajputmanish3101@gmail.com

Dr. Raju Ranjan

School of Computer Science And
Engineering
Galgotias University
Greater Noida, Uttar Pradesh, India
Drraju.ranjan@galgotiasuniversity.edu.in

Abstract— Nowadays social media has become one of the important or popular platforms for sharing news. The easy access to everyone with no cost to low cost and instant circulation of the information/fake news led people to look for and consume news from social media. And we know that everything has advantages along with that disadvantages is also there. so, in the same way, spreading fake news on social media becomes the most significant problem nowadays. And in this research paper, we will detect whether the news is fake or true with the help of machine learning algorithms.

In this research paper, we use three classification models that are logistic regression, decision tree classification, and gradient boosting classification to detect news whether it is fake or true. And after detection, we also show a comparison between these classification methods. Which method or we can say model gives the higher accuracy to predict the news?

In this research paper, we will extend our research from the current level of research to the next level by doing further research on recommended works in the previous research paper.

Keywords— fake news, detection, comparison, logistic regression, Decision tree classification, Gradient Boosting classification.

I. INTRODUCTION

Even though social media has only been around for a little over a decade, we are already noticing its influence on our daily lives. From Twitter to Facebook to Instagram[1], these platforms make it easy for users to share information with their followers in a matter of seconds. However, when you scroll through your feed, you might start noticing headlines that seem fishy[2]. Whether they promote fear or false hope, these sources are sometimes the most convincing because they look like they come from people you know and trust. The impact of fake news is very bad not only on our society but also impacts the decision that we take in our life. When wrong/fake information or misinformation is spread through media, Somehow, It declines the trust in media. The major negative effect of this is, It increases division and polarization.

Fake news stories[3] are often designed to appeal to our confirmation biases, and they can be difficult to distinguish from accurate reporting. This is compounded by the fact that many people get their news from social media platforms, which are not always reliable sources of information.

The effects of fake news go beyond simply causing confusion or misinformation[4]. Fake news stories can have real-world consequences, such as inciting violence or spreading false rumors about businesses or individuals. Politicians also use to spread fake news to the young generation to impact their vote power in the upcoming election.

In fact, in today's world, Fake news is spreading fast to the young generation when it comes to any election because election parties want to influence the young generation's mind to win the election on any condition.

Government is also very serious about spreading fake news and its impact on the public. Numerous Machine Learning models have been proposed to detect fake news and each has its advantages and disadvantages.

In this research paper, we are proposing some research to detect fake news using Machine Learning.

II. LITERATURE SURVEY

The literature that is now available has described a variety of automatic methods for spotting deceptive posts and fake news. Since there are many other ways to identify fake news, like employing clickbait to spread rumours and chatbots to spread misinformation. Twitter and Different clickbait[5] are available on other social media networks to promote sharing and liking of fake news. The findings suggest that 15% of tweets are fraudulent, 45% are genuine, and the remaining messages are unclear. The previous research paper spotted 6 different Machine Learning algorithms for detecting fake news online.

In that research paper, data is split into two categories. One is the training dataset and another is the testing dataset. When we talk about previous research paper evaluation metrics, they have used precision, recall, f1 score, and ROC - AUC. Metrics.

Machine Learning algorithms play an important role in detecting Fake News online. In the previous dataset, They used these ML algorithms to predict fake news online. These datasets were collected from various sources. These sources are included YouTube, Facebook, Twitter, and Online Blogs.

Previous research papers included 3 types of datasets that are easily and freely available on the internet. Performing different



ML algorithms on these datasets gave us different results. One of these datasets applying one ML algorithm on it gives 94.5% accuracy overall for fake news detection. There are more than 10 research papers available on fake news detection with good accuracy.

We have studied another research paper where they have used Deep Learning Algorithms to detect fake news online. That research paper used various ML algorithms such as CNN, Deep CNN, and CNN+LSTM. Their overall accuracy is 98.90%. This accuracy is quite better than the previous research paper about fake news detection online.

A. Type of Data in Postings on Social Media

There are three major ways by which social media networking sites can be used.

Text - Computational linguistics, which focuses on the creation of text systematically and semantically, analyses text (multilingually). Since texts make up the majority of posts[6], extensive effort has been done into their analysis.

Multimedia - A single post incorporates various media types. Audio, video, photos, and graphics may all be included. Without the viewers even caring about the text, this is quite appealing and grabs their attention.

Hyperlinks - By allowing the author of the piece to cross-reference to many sources, Hyperlinks enhance viewers' confidence. preventing Fake news has existed for a long time. Since it first emerged, this issue has been addressed because of the detrimental effects it had on both the technological and political premises. To face or somehow manage false news[7] – The spreading of fake news through online and offline mediums is not a recent trend.

This is one of the major problems that India and the world currently going on and, Due to this, it impacts various technical and political spheres negatively which can be very harmful to a nation. It must be addressed since, as a result of technology advancements, people's daily reliance on social networking sites for news is on the rise, and has no intention of going away any time soon.

To identify fake news, we offered features in the preceding section that were taken from many sources, such as news content and social context. For numerous current methodologies, we go into detail in this section on the model construction process. We categorize current techniques specifically.

III. MACHINE LEARNING MODEL

A. Supervised Machine Learning

A form of supervised machine learning is one in which the dependent variable is known and based on known data, we train the model and predict the output. And in supervised machine learning, we have a labeled dataset.

In this research paper we used three supervised machine learning methods to predict the news whether it is fake or true, Let's discuss them one by one. unavoidable.

1) Logistic Regression

Logistic regression predicts outputs in the form of a linear equation based on the input. It is a classification technique for binary responses that are designed to assess which independent variables are related to the outcome of interest. Besides logistic regression being limited to categorical values, it also focuses on modeling relationships between two variables[5].

Logistic regression is applied to fields such as marketing, economics, and psychology for making predictions about human behavior.

Result

When we use logistic regression to predict fake news, we observed an accuracy of around 98% as shown in the above diagram.

In this logistic regression, We have used a dataset that we have collected from sources like YouTube, Online websites, Blogs, and Kaggle.

Logistic Regression

```
from sklearn.linear_model import LogisticRegression
```

```
LR = LogisticRegression()  
LR.fit(xv_train,y_train)
```

```
LogisticRegression()
```

```
LR.score(xv_test,y_test)
```

```
0.9858288770053476
```

```
Pred_LR = LR.predict(xv_test)
```

```
Print(classification_report(y_test,pred_LR))
```

	precision	recall	f1-score	support
0	0.99	0.99	0.99	5899
1	0.98	0.99	0.99	5321
accuracy			0.99	11220
macro avg	0.99	0.99	0.99	11220
weighted avg	0.99	0.99	0.99	11220

2) Decision Tree Classification: -

A collection of input data may be transformed into a set of discrete class labels using the decision tree classification technique. Decision Tree Classification[6] takes the form of a tree: each branch and node correspond to a different variable and its level in what's called the "tree hierarchy".



Example:

Consider a sample of 30 students with the following three variables: height, class (IX or X), and gender (boy or girl) (5 to 6 ft). Out of these 30, 15 enjoy playing cricket. I now want to develop a model to anticipate who will play cricket during their free time. To solve this issue/problem, we must separate the students who play cricket as a hobby depending[7] on the most important input variable among the three.

This is how the Decision tree algorithm is supposed to work.:

Decision Tree

```
from sklearn.linear_model import DecisionTreeClassifier

DT = DecisionTreeClassifier()
DT.fit(xv_train,y_train)

DecisionTreeClassifier()

DT.score(xv_test,y_test)

0.9973262032085561

Pred_DT = DT.predict(xv_test)

Print(classification_report(y_test,pred_DT))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	5899
1	1.00	1.00	1.00	5321
accuracy			1.00	11220
macro avg	1.00	1.00	1.00	11220
weighted avg	1.00	1.00	1.00	11220

Result

When we use a decision tree to predict fake news, we observed an accuracy of around 99.73% as shown in the above diagram.

The dataset which we have used in previous ML algorithms to predict Fake News detection is the same one used in this algorithm for predicting fake news.

3) Gradient Boosting Classification

Gradient boosting is a technique in machine learning that's designed to address the problem of weak learning in predictive models. Gradient boosting[8] works by reweighting training examples so that those with the smallest loss function values contribute more to the model. There are several variations of GBDT algorithms[9] that deal with specific tasks in different ways[10].

Gradient Boosting classifier

```
from sklearn.linear_model import GradientBoostingClassifier

GBC = GradientBoostingClassifier(random_state=0)
GBC.fit(xv_train,y_train)

GradientBoostingClassifier(random_state=0)

GBC.score(xv_test,y_test)

0.9959001782531194

Pred_GBC = GBC.predict(xv_test)

Print(classification_report(y_test,pred_GBC))
```

	precision	recall	f1-score	support
0	1.00	0.99	1.00	5899
1	0.99	1.00	1.00	5321
accuracy			1.00	11220
macro avg	1.00	1.00	1.00	11220
weighted avg	1.00	1.00	1.00	11220

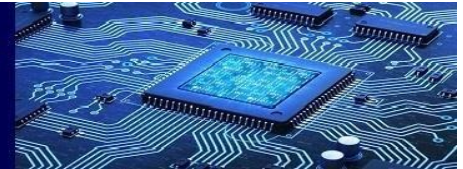
Result

When we use a decision tree to predict fake news, we observed an accuracy of around 99.59% as shown in the above diagram.

IV. CONCLUSION

In this research paper, we detect fake news using three classification methods that are logistic regression, decision tree classification, and gradient boosting classification, and find the accuracy of all the models on the data. And find out that the decision tree model gives the best result with 99.73% accuracy.

We have performed these ML algorithms analysis on various types of datasets that we have collected through many platforms like Youtube, Google, etc.



Algorithms	Formulas	Accuracy
Logistic Regression	$\frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}}$	98%
Decision Tree	$\text{Entropy} = \sum_{i=1}^n P_i * \log(P_i)$	99.73%
Gradient Boosting Classifier	$F_m(x) = F_{m-1}(x) + \gamma_{\text{optimum}} h_m(x)$	99.59%

[17] Mykhailo Granik and Volodymyr Mesyura. "Fake news detection using naive Bayes classifier." First Ukraine Conference on Electrical and Computer Engineering (UKRCON). Ukraine: IEEE. 2017

REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955. (references)
- [2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [4] K. Elissa, "Title of paper if known," unpublished.
- [5] Manzoor, S. I., & Singla, J. (2019, April). Fake news detection using machine learning approaches: A systematic review. In 2019 3rd international conference on trends in electronics and informatics (ICOEI) (pp. 230-234)
- [6] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1), 22-36
- [7] R. Nicole, "Title of paper with the only first word capitalized," *J. Name Stand. Abbrev.*, in press.
- [8] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Trans. J. Magn. Japan*, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetism Japan, p. 301, 1982].
- [9] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
- [10] Bovet, Alexandre, and Hernán A. Makse. "Influence of fake news in Twitter during the 2016 US presidential election." *Nature communications* 10.1 (2019): 1-14. The science of fake news.
- [11] <https://science.sciencemag.org/content/359/6380/1094>. *Summary Science* 09 Mar 2018: Vol. 359,
- [12] Huang, T.-Q. (n.d.) https://www.researchgate.net/figure/Pseudo-code-of-information-gain-based-recursive-feature-elimination-procedure-with-SVM_fig2_228366941 2018
- [13] Researchgate.net. Available at: Available at: https://www.researchgate.net/figure/Pseudocode-of-naive-bayes-algorithm_fig2_325937073. 2018.
- [14] Researchgate.net. Available at: https://www.researchgate.net/figure/Pseudocode-for-KNN-classification_fig7_260397165, 2014.
- [15] Economic and Social Research Council. *Using Social Media*. Available at: <https://esrc.ukri.org/research/impact-toolkit/social-media/using-social-media>
- [16] Gil, P. Available at: <https://www.lifewire.com/what-exactly-is-twitter-2483331>. 2019, April 22.