ISSN No : 2454-4221(Print)
ISSN No : 2454-423X(Online)

International Journal of Research in Advanced Computer Science Engineering

A Peer Reviewed Open Access International Journal

www.ijracse.com

Fake news Detection Using Ensemble Learning Techniques

S. Mahlakshmi[1], B. Bharadwaj[2], M. Durga Avinash[3], K. Dinesh Kumar[4], M.Ratna Mohitha[5]*

Department of Computer Science and Engennering[12345],

Aditya Institute of Technology and Management, India, Andhra Pradesh, Tekkali 532201.

Corresponding Author: r.mohithamaddula@gmail.com

Abstract:

False information presented as news is referred to as fake news. Fake news is a major issue that threatens to destroy people's reputations, puts many different businesses and governmental agencies in peril, and has terrible negative effects on individuals and mankind. So, it's important to distinguish between legitimate and false news. Which is very difficult through manually. Thus, a machine learning model is required to identify whether the news is false or not. In this work, we employed Ensemble machine learning models, such as Extra tree Classifier, gradient boost, extreme gradient boost, and light gradient boost. We also find the model with the best level of accuracy. In order to assess if the news was accurate or fake, manual testing was also done.

Keywords: Extra tree Classifier, Gradient Boost, Extreme Gradient Boost, Light Gradient Boost, Fake, Real, Manual testing.

1. Introduction

In today's digital environment, fake news is one of the most pressing issues since it spreads quickly via social media and internet platforms. "Fabricated material that replicates news media content in form but differs in organizational method or goal" is what is meant by fake news. Untrue news serious problems in this contemporary environment. False news poses a danger to the worlds of politics, business, finance, education, and democracy. Public opinion could alter, and one might get a mistaken impression. Researchers and data scientists have created a range of methods and frameworks for detecting fake news in order to solve this issue, including ones that concentrate on recognizing false news items based on their source or content. Propaganda, gossip, disinformation, hoaxes, satire, clickbait, misinformation, and junk news are just a few examples of fake news' many varieties. False connections, misleading content, false context, imposter content, manipulated content, and fabricated content are a few examples of several sorts of fake news. Real news' effect might be diminished by fake news. So, this type of unethical behaviour must be controlled. This study features are listed here:

- Our approach makes it simple to recognize false information.

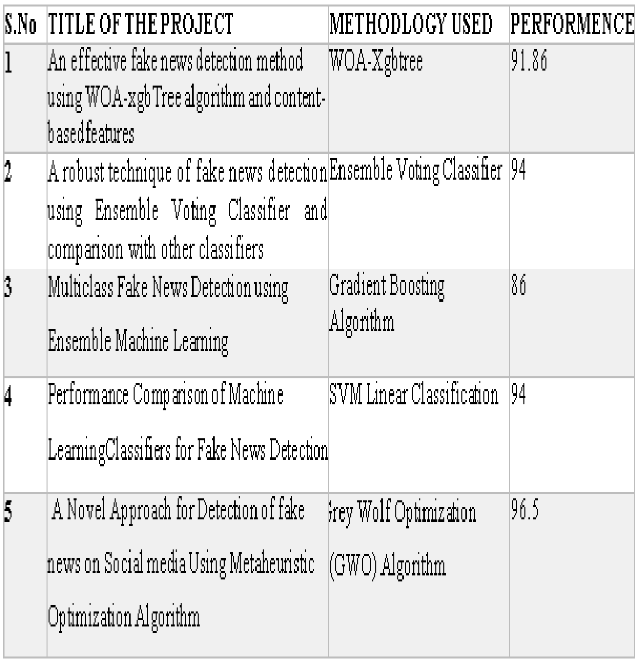
- Fake news that harms a person's reputation can be stopped.

- We can control the unethical behaviour.

Cite this article as: S.Mahlakshmi, B.Bharadwaj, M.Durga Avinash, K.Dinesh Kumar & M.RatnaMohitha, "Fake news Detection Using Ensemble Learning Techniques", International Journal of Research in Advanced Computer Science Engineering, (IJRACSE), Volume 8 Issue 11, April 2023, Page 61-67.

ISSN No : 2454-4221(Print)
ISSN No : 2454-423X(Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

2. Literature Survey



S.No	TITLE OF THE PROJECT	METHODLGY USED	PERFORMENCE
1	An effective fake news detection method using WOA-xgbTree algorithm and content-basedfeatures	WOA-Xgbtree	91.86
2	A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers	Ensemble Voting Classifier	94
3	Multiclass Fake News Detection using Ensemble Machine Learning	Gradient Boosting Algorithm	86
4	Performance Comparison of Machine LearningClassifiers for Fake News Detection	SVM Linear Classification	94
5	A Novel Approach for Detection of fake news on Social media Using Metaheuristic Optimization Algorithm	Grey Wolf Optimization (GWO) Algorithm	96.5

3. Methodology

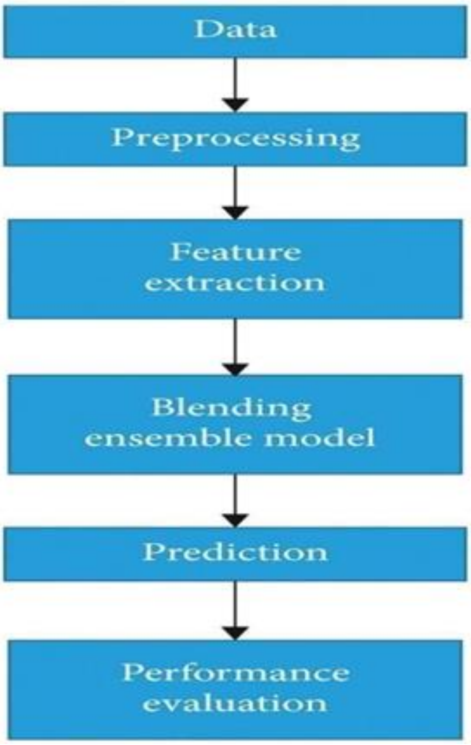


Figure-1: Flow of Action.

3.1 Dataset

On the Kaggle platform, we gathered a dataset that includes two datasets: real news and fake news. The genuine dataset contains all of the true news stories, which are divided into 21418 rows and 4 columns (Title, Text, Subject matter, Date), whereas the fake dataset contains all of the false news stories, which are divided into 23501 rows and 4 columns (Tittle, Text, Subject, Date) To determine if news is false or not, we integrate then two datasets and do manual testing.

SI. No.	Dataset Name	Sources	Attribute Type	Number of Attributes	Number of Instances
1	True Dataset	Kaggle Repository	Text	4	21418
2	Fake Dataset	Kaggle Repository	Text	4	23501

Table-2 Dataset Information.

3.2 Preprocessing

An essential phase in the data mining process is data pre-processing. It describes the processes of preparing data for analysis by cleansing, converting, and integrating it. The purpose of data preprocessing is to enhance the data's quality and suitability for the particular data mining operation.

• Data cleaning: in this process, missing, inconsistent, or unnecessary data are found and removed. This might involve eliminating redundant data, adding values when they are missing, and dealing with outliers.

• Data integration: at this stage, data from many sources, including databases, spreadsheets, and text files, are combined. One consistent picture of the data is what integration aims to produce.

• Data transformation: in this process, the data is transformed into a format that is more suited for data mining. This may involve encoding category data, producing dummy variables, and normalizing numerical data.

• Data reduction: a subset of the data that is pertinent to the data mining objective is chosen using this process. This may entail feature extraction or feature selection (choosing a subset of the variables) (extracting new variables from the data).

• Data discretization: this process turns continuous numerical data into categorical data that may be utilized in categorical data mining techniques like decision trees. These procedures increase the effectiveness of data mining and improve the precision of the findings.

3.3 Algorithms

Due to the several algorithms used to determine if news is false or not. Nonetheless, we select the algorithm that provides the highest level of accuracy compared to the earlier models, thus we select ensemble learning models[2][3] such as Extra Tree Classifier, Gradient Boost, Extreme Gradient Boost[1], and Light Gradient Boost.

Ensemble learning[2][3]: Machine learning ensemble approaches integrate the insights from several learning models to enable more precise and better conclusions. The weak learners unite in assembling strategies to create a strong model.

Extra Tree Classifier:

Very Randomized Trees Classifier, also known as Extra Trees Classifier, is a form of ensemble learning[2][3]approach that combines the findings of several de-correlated decision trees gathered in a "forest" to produce its classification outcome.

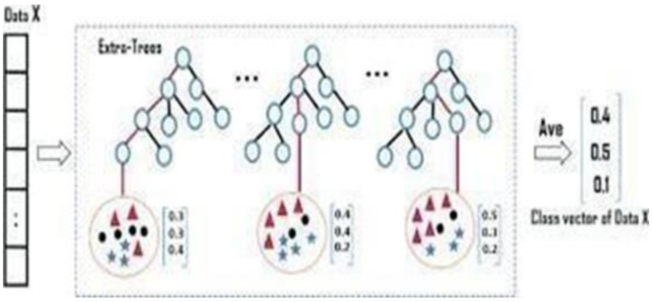


Figure-2: Working of Extra Tree Classifier.

Gradient Boost:

The basic idea behind boosting methods is that after creating a model using the training dataset, we create a second model to fix any mistakes in the original one.

This algorithm's fundamental principle is to create models consecutively while attempting to minimize the mistakes of the prior model.

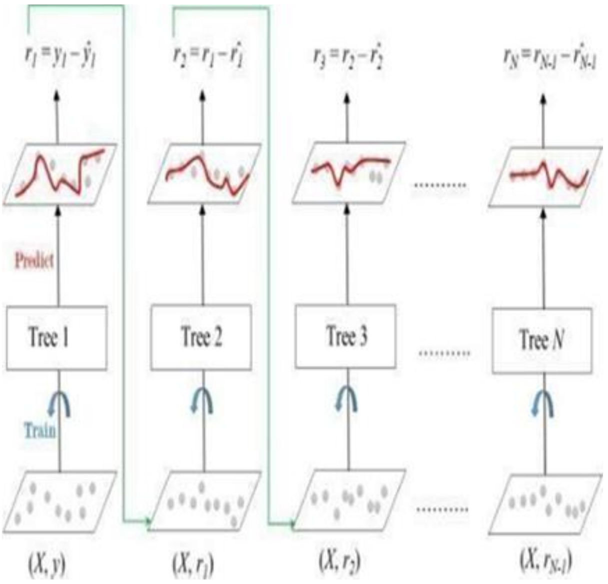
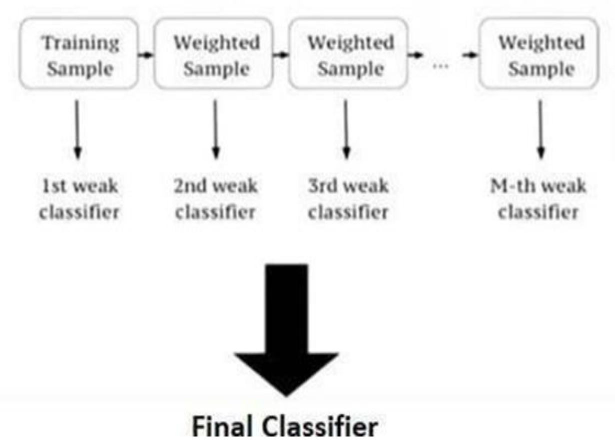


Figure-3: Working of Gradient Boost.

ExtremeGradient Boost:

Decision trees are generated sequentially in this approach. It is a kind of software library that was primarily created to increase model performance and speed. Weights are significant in XGBoost[1]. Each independent variable is given a weight before being put into the decision tree that forecasts outcomes. The variables are subsequently put into the second decision tree with an enhanced weight for variables that the tree incorrectly anticipated. These distinct classifiers/predictors are then combined to produce a robust and accurate model.



LightGradient Boost:

LightGBM is a gradient boosting framework built on decision trees that improves model performance by consuming fewer memory. It satisfies the constraints using two unique techniques: gradient-based one side sampling and exclusive feature bundling (EFB). Light gradient increase offered a number of benefits such as greater training effectiveness and quickness, use less memory, more precision, support for distributed, parallel, and GPU learning, ability to manage massive amounts of data.

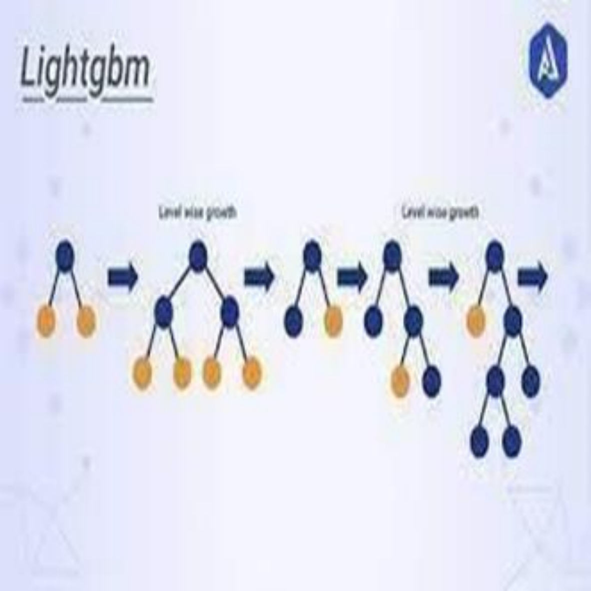
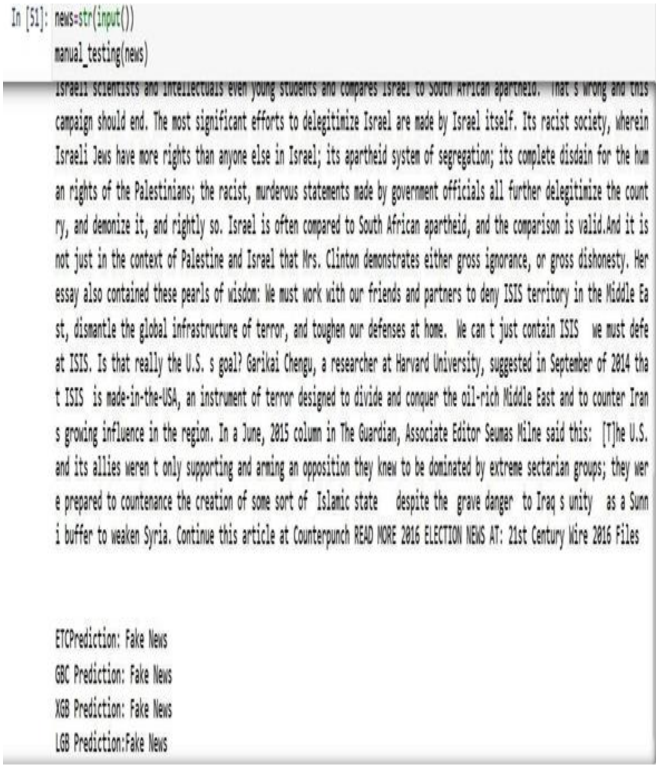


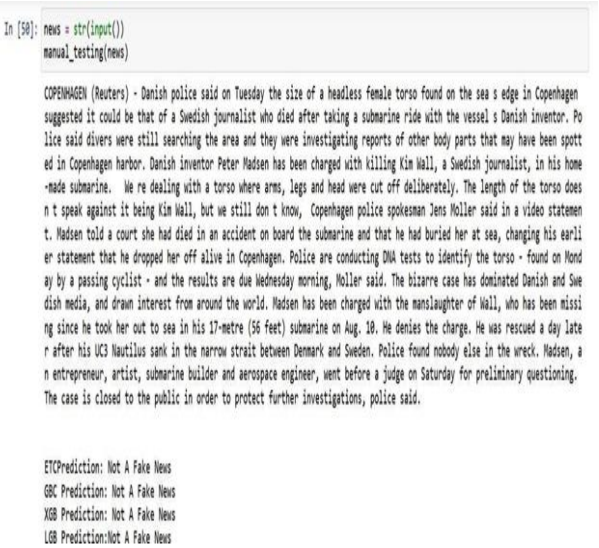
Figure-3: Working of LightGradient Boost.

4. Results:



```
In [51]: news=str(input())
         manual_testing(news)

         Israeli scientists and intellectuals even young students and compares Israel to South African apartheid.  That s wrong and this
         campaign should end. The most significant efforts to delegitimize Israel are made by Israel itself. Its racist society, wherein
         Israeli Jews have more rights than anyone else in Israel; its apartheid system of segregation; its complete disdain for the hum
         an rights of the Palestinians; the racist, murderous statements made by government officials all further delegitimize the count
         ry, and demonize it, and rightly so. Israel is often compared to South African apartheid, and the comparison is valid.And it is
         not just in the context of Palestine and Israel that Mrs. Clinton demonstrates either gross ignorance, or gross dishonesty. Her
         essay also contained these pearls of wisdom: We must work with our friends and partners to deny ISIS territory in the Middle Ea
         st, dismantle the global infrastructure of terror, and toughen our defenses at home.  We can t just contain ISIS   we must defe
         at ISIS. Is that really the U.S. s goal? Garikai Chengu, a researcher at Harvard University, suggested in September of 2014 tha
         t ISIS  is made-in-the-USA, an instrument of terror designed to divide and conquer the oil-rich Middle East and to counter Iran
         s growing influence in the region. In a June, 2015 column in The Guardian, Associate Editor Seumas Milne said this:  [T]he U.S.
         and its allies weren t only supporting and arming an opposition they knew to be dominated by extreme sectarian groups; they wer
         e prepared to countenance the creation of some sort of  Islamic state   despite the  grave danger  to Iraq s unity   as a Sunn
         i buffer to weaken Syria. Continue this article at Counterpunch READ MORE 2016 ELECTION NEWS AT: 21st Century Wire 2016 Files


         ETCPrediction: Fake News
         GBC Prediction: Fake News
         XGB Prediction: Fake News
         LGB Prediction:Fake News
```

The above News is predicted as fake news.

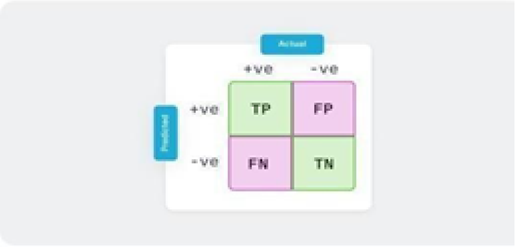


```
In [50]: news = str(input())
         manual_testing(news)

         COPENHAGEN (Reuters) - Danish police said on Tuesday the size of a headless female torso found on the sea s edge in Copenhagen
         suggested it could be that of a Swedish journalist who died after taking a submarine ride with the vessel s Danish inventor. Po
         lice said divers were still searching the area and they were investigating reports of other body parts that may have been spott
         ed in Copenhagen harbor. Danish inventor Peter Madsen has been charged with killing Kim Wall, a Swedish journalist, in his home
         -made submarine.   We re dealing with a torso where arms, legs and head were cut off deliberately. The length of the torso does
         n t speak against it being Kim Wall, but we still don t know,  Copenhagen police spokesman Jens Moller said in a video statemen
         t. Madsen told a court she had died in an accident on board the submarine and that he had buried her at sea, changing his earli
         er statement that he dropped her off alive in Copenhagen. Police are conducting DNA tests to identify the torso - found on Mond
         ay by a passing cyclist - and the results are due Wednesday morning, Moller said. The bizarre case has dominated Danish and Swe
         dish media, and drawn interest from around the world. Madsen has been charged with the manslaughter of Wall, who has been missi
         ng since he took her out to sea in his 17-metre (56 feet) submarine on Aug. 10. He denies the charge. He was rescued a day late
         r after his UC3 Nautilus sank in the narrow strait between Denmark and Sweden. Police found nobody else in the wreck. Madsen, a
         n entrepreneur, artist, submarine builder and aerospace engineer, went before a judge on Saturday for preliminary questioning.
         The case is closed to the public in order to protect further investigations, police said.


         ETCPrediction: Not A Fake News
         GBC Prediction: Not A Fake News
         XGB Prediction: Not A Fake News
         LGB Prediction:Not A Fake News
```

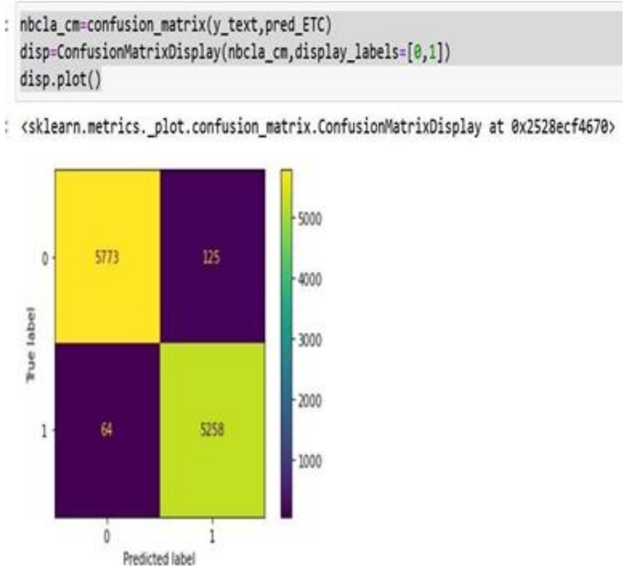
The above News is predicted as not a fake news.

4.1 Confusion Matrix[4]:

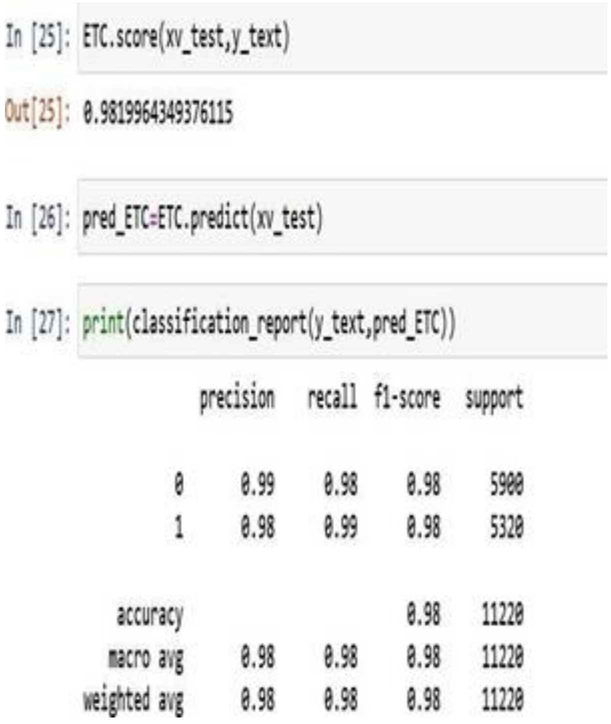
A table called a confusion matrix is used to describe how well a classification system performs. The potential of a classifier may be accurately assessed using a confusion matrix. Each diagonal element represents a successfully categorized result. The off diagonals of the confusion matrix show the misclassified results. he four quadrants are defined as True Negative (TN), True Positive (TP), False Positive (FP), False Negative (FN).



Accuracy = (True positives + True Negatives)/ (True positives + True negatives + False positives + False negatives)
Precision = TP / (TP + FP) (where TP is True Positives and FP is False Positives)
Recall = TP / (TP + FN) (where FN is False Negative) F1 score = (precision * recall) / (precision + recall)

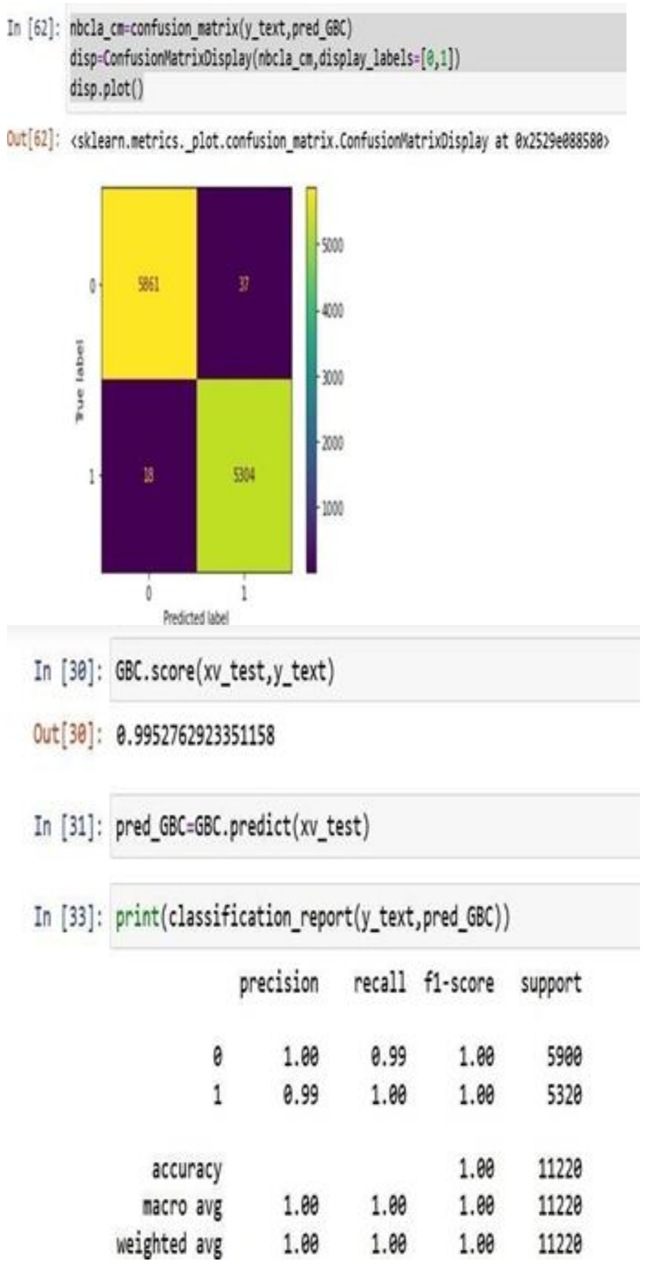


```
: nbcla_cm=confusion_matrix(y_text,pred_ETC)
  disp=ConfusionMatrixDisplay(nbcla_cm,display_labels=[0,1])
  disp.plot()

: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x2528ecf4670>
```



```
In [25]: ETC.score(xv_test,y_text)

Out[25]: 0.9819964349376115

In [26]: pred_ETC=ETC.predict(xv_test)

In [27]: print(classification_report(y_text,pred_ETC))

              precision    recall  f1-score   support

           0       0.99      0.98      0.98      5900
           1       0.98      0.99      0.98      5320

    accuracy                           0.98     11220
   macro avg       0.98      0.98      0.98     11220
weighted avg       0.98      0.98      0.98     11220
```

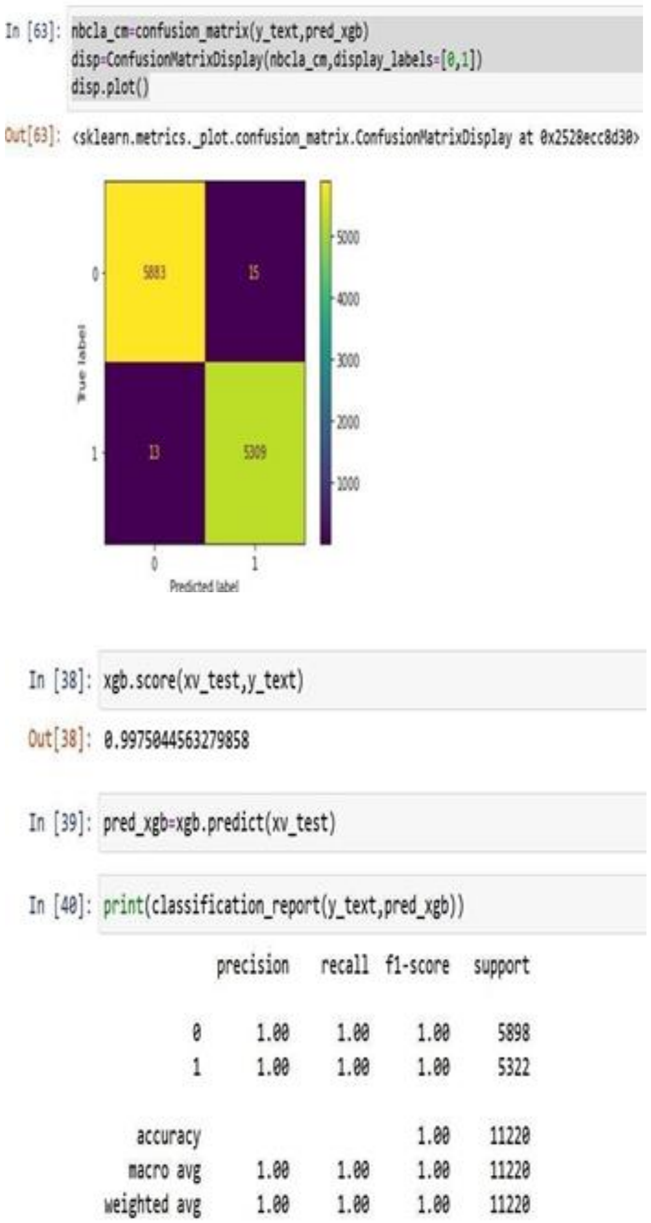
(a)



```
In [62]: nbcla_cm=confusion_matrix(y_text,pred_GBC)
         disp=ConfusionMatrixDisplay(nbcla_cm,display_labels=[0,1])
         disp.plot()

Out[62]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x2529e088580>
```

```
In [30]: GBC.score(xv_test,y_text)

Out[30]: 0.9952762923351158

In [31]: pred_GBC=GBC.predict(xv_test)

In [33]: print(classification_report(y_text,pred_GBC))

              precision    recall  f1-score   support

           0       1.00      0.99      1.00      5900
           1       0.99      1.00      1.00      5320

    accuracy                           1.00     11220
   macro avg       1.00      1.00      1.00     11220
weighted avg       1.00      1.00      1.00     11220
```

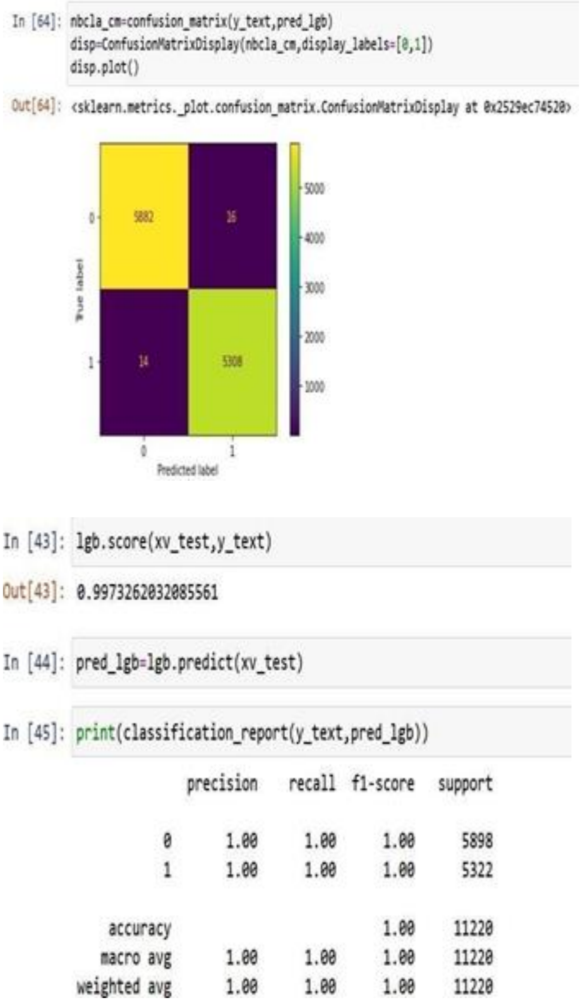
(b)



```
In [63]: nbcla_cm=confusion_matrix(y_text,pred_xgb)
         disp=ConfusionMatrixDisplay(nbcla_cm,display_labels=[0,1])
         disp.plot()

Out[63]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x2528ecc8d30>
```

```
In [38]: xgb.score(xv_test,y_text)

Out[38]: 0.9975044563279858

In [39]: pred_xgb=xgb.predict(xv_test)

In [40]: print(classification_report(y_text,pred_xgb))

              precision    recall  f1-score   support

           0       1.00      1.00      1.00      5898
           1       1.00      1.00      1.00      5322

    accuracy                           1.00     11220
   macro avg       1.00      1.00      1.00     11220
weighted avg       1.00      1.00      1.00     11220
```

(c)



```
In [64]: nbcla_cm=confusion_matrix(y_text,pred_lgb)
         disp=ConfusionMatrixDisplay(nbcla_cm,display_labels=[0,1])
         disp.plot()

Out[64]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x2529ec74520>
```

```
In [43]: lgb.score(xv_test,y_text)

Out[43]: 0.9973262032085561

In [44]: pred_lgb=lgb.predict(xv_test)

In [45]: print(classification_report(y_text,pred_lgb))

              precision    recall  f1-score   support

           0       1.00      1.00      1.00      5898
           1       1.00      1.00      1.00      5322

    accuracy                           1.00     11220
   macro avg       1.00      1.00      1.00     11220
weighted avg       1.00      1.00      1.00     11220
```

(d)

a) Confusion matrix for Extra Tree Classifier[4].
b) Confusion matrix for Gradient Boost[4].
c) Confusion matrix for Extreme Gradient Boost[1][4].
d) Confusion matrix for Light Gradient Boost[4].

5. Conclusion and Discussion

In order to manually classify whether the news is false or true, we mixed part of the news from the true and false databases. We examined the accuracy of the Gradient Boost, Extreme Gradient Boost[1], and Light Gradient Boost algorithms[4]. Thus, we can say that Light Gradient Boost had the highest accuracy. Lastly, we categorize whether a piece of news is fake or not. If the news is false, the outcome states as fake otherwise, it states as Not fake news.

6. References

1) Saeid Sheikhi, An effective fake news detection method using WOA-xgbTree algorithm and content- based features, Applied Soft Computing 109 (2021) 107559, 2021.
2) Atik Mahabub, A robust technique of fake news detection using Ensemble Voting Classifer and comparison with other classifers, SN Applied Sciences (2020) 2:525, 2020.
3) Rohit Kumar Kaliyar, Anurag Goswami, Pratik Narang, Multiclass Fake News Detection using Ensemble Machine Learning, IEEE 978-1-7281-4392-7/19, 2019.
4) Smitha. N, Bharath .R, Performance Comparison of Machine Learning Classifiers for Fake News Detection, Proceedings of the Second International Conference on Inventive Research in Computing Applications (ICIRCA-2020) IEEE Xplore Part Number: CFP20N67-ART; ISBN: 978-1-7281-5374-2, 2020.
5) Feyza Altunbey Ozbay, Bilal Alatas, A Novel Approach for Detection of Fake News on Social
Media Using Metaheuristic Optimization Algorithms, ELEKTRONIKA IR ELEKTROTECHNIKA, ISSN 1392-1215, VOL. 25, NO. 4, 201