```
!kaggle datasets download -d "ahmedsmara/full-fake-news"
```

```
Dataset URL: https://www.kaggle.com/datasets/ahmedsmara/full-fake-news
License(s): unknown
Downloading full-fake-news.zip to /content
100% 37.7M/37.7M [00:03<00:00, 18.7MB/s]
100% 37.7M/37.7M [00:03<00:00, 11.5MB/s]
```

```
!unzip /content/full-fake-news.zip
```

```
Archive:  /content/full-fake-news.zip
  inflating: full dataset.csv
```

```
import pandas as pd
df=pd.read_csv("/content/full-fake-news.zip")
df
```

	title	author	text	label
0	a news release	federation-american-immigration-reform	Unemployment has been on the rise throughout W...	1
1	Black Turnout Soft in Early Voting, Boding Ill...	Henry Wolff	Black Turnout Soft in Early Voting, Boding Ill...	1
2	a television interview	chris-abele	Says Milwaukee County buses are no less safe n...	1
3	10 Things I Learned From Being My Own General ...	Luke Stranahan	One of the foundations of living a good life i...	1
4	a statement responding to Gov. Rick Scott's St...	rod-smith	Says Rick Scotts proposed budget would lay off...	0
...	...	...	...	...
30933	a TV ad.	bill-richardson	Bill Richardson passed the first law in the na...	0
30934	Khodorkovsky: Putin is not going to cozy up to...	NaN	Print version Font Size The number of predicti...	1
30935	The Only Coup Is a Stolen Election	Dave Hodges	\n\nWars and rumor of wars. There is no coup. ...	1
30936	Doomed Jet Carrying Brazilian Team Reportedly ...	Nicholas Casey and Ewan MacKenna	The pilot of a plane that crashed late Monday ...	0
30937	Burger King 'O.K. Google' Ad Doesn't Seem O.K....	Sapna Maheshwari	Google appeared to stymie a marketing stunt on...	0

30938 rows × 4 columns

```
df.shape
```

```
(30938, 4)
```

```
df.columns
```

```
Index(['title', 'author', 'text', 'label'], dtype='object')
```

```
df.head()
```

	title	author	text	label
0	a news release	federation-american-immigration-reform	Unemployment has been on the rise throughout W...	1
1	Black Turnout Soft in Early Voting, Boding Ill...	Henry Wolff	Black Turnout Soft in Early Voting, Boding Ill...	1
2	a television interview	chris-abele	Says Milwaukee County buses are no less safe n...	1
3	10 Things I Learned From Being My Own General ...	Luke Stranahan	One of the foundations of living a good life i...	1
4	a statement responding to Gov. Rick Scott's St...	rod-smith	Says Rick Scotts proposed budget would lay off...	0

```
df.tail()
```

	title	author	text	label
30933	a TV ad.	bill-richardson	Bill Richardson passed the first law in the na...	0
30934	Khodorkovsky: Putin is not going to cozy up to...	NaN	Print version Font Size The number of predicti...	1
30935	The Only Coup Is a Stolen Election	Dave Hodges	\n\nWars and rumor of wars. There is no coup. ...	1
30936	Doomed Jet Carrying Brazilian Team Reportedly ...	Nicholas Casey and Ewan MacKenna	The pilot of a plane that crashed late Monday ...	0
30937	Burger King 'O.K. Google' Ad Doesn't Seem O.K....	Sapna Maheshwari	Google appeared to stymie a marketing stunt on...	0

```
df.isna().sum()
```

	0
title	558
author	1957
text	39
label	0

dtype: int64

```
df['title'].fillna("No Title Available", inplace=True)
```

```
<ipython-input-10-d262fdf179f7>:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace metho
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a cop

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to pe
```

```
df['title'].fillna("No Title Available", inplace=True)
```

```
df['author'].fillna("Unknown Author", inplace=True)
```

> `<ipython-input-11-a284e3d6364b>:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method`
> `The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a cop`
>
> `For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to pe`
>
>   `df['author'].fillna("Unknown Author", inplace=True)`

```
df['text'].fillna("No Text Available", inplace=True)
```

> `<ipython-input-12-1203a93882cb>:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace metho`
> `The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a cop`
>
> `For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to pe`
>
>   `df['text'].fillna("No Text Available", inplace=True)`

```
df.isna().sum()
```

	0
title	0
author	0
text	0
label	0

dtype: int64

```
df.shape
```

> (30938, 4)

```
df.duplicated().sum()
```

> 110

```
df.drop_duplicates(inplace=True)
```

```
df.shape
```

> (30828, 4)

```
df['author'].value_counts()
```

	count
author	
Unknown Author	1932
barack-obama	481
donald-trump	273
Pam Key	243
hillary-clinton	235
...	...
markos-moulitsas	1
Julie Hirschfeld Davis, Thomas Kaplan and Robert Pear	1
end-citizens-united	1
moms-rising	1
Peter Baker and Ron Nixon	1

7099 rows × 1 columns

dtype: int64

```
df['label'].value_counts()
```

	count
label	
0	16084
1	14744

dtype: int64

```
df['text'].value_counts()
```

	count
text	
No Text Available	73
Leave a Reply Click here to get more info on formatting (1) Leave the name field empty if you want to post as Anonymous. It's preferable that you choose a name so it becomes clear who said what. E-mail address is not mandatory either. The website automatically checks for spam. Please refer to our moderation policies for more details. We check to make sure that no comment is mistakenly marked as spam. This takes time and effort, so please be patient until your comment appears. Thanks. (2) 10 replies to a comment are the maximum. (3) Here are formating examples which you can use in your writing:\bold text\ results in bold text \<i>italic text\</i> results in italic text (You can also combine two formating tags with each other, for example to get bold-italic text.)\emphasized text\ results in emphasized text \strong text\ results in strong text \<q>a quote text\</q> results in a quote text (quotation marks are added automatically) \<cite>a phrase or a block of text that needs to be cited\</cite> results in: a phrase or a block of text that needs to be cited \<blockquote>a heavier version of quoting a block of text...\</blockquote> results in: a heavier version of quoting a block of text that can span several lines. Use these possibilities appropriately. They are meant to help you create and follow the discussions in a better way. They can assist in grasping the content value of a comment more quickly. and last but not least:\Name of your link\ results in Name of your link (4) No need to use this special character in between paragraphs: ; You do not need it anymore. Just write as you like and your paragraphs will be separated. The "Live Preview" appears automatically when you start typing below the text area and it will show you how your comment will look like before you send it. (5) If you now think that this is too confusing then just ignore the code above and write as you like. Name:	39
	28
Click Here To Learn More About Alexandra's Personalized Essences Psychic Protection Click Here for More Information on Psychic Protection! Implant Removal Series Click here to listen to the IRP and SA/DNA Process Read The Testimonials Click Here To Read What Others Are Experiencing! Copyright © 2012 by Galactic Connection. All Rights Reserved. \nExcerpts may be used, provided that full and clear credit is given to Alexandra Meadors and www.galacticconnection.com with appropriate and specific direction to the original content. Unauthorized use and/or duplication of any material on this website without express and written permission from its author and owner is strictly prohibited. Thank you. \nPrivacy Policy \nBy subscribing to GalacticConnection.com you acknowledge that your name and e-mail address will be added to our database. As with all other personal information, only working affiliates of GalacticConnection.com have access to this data. We do not give GalacticConnection.com addresses to outside companies, nor will we ever rent or sell your email address. Any e-mail you send to GalacticConnection.com is completely confidential. Therefore, we will not add your name to our e-mail list without your permission. Continue reading... Galactic Connection 2016	Design & Development by AA at Superluminal Systems Sign Up forOur Newsletter \nJoin our newsletter to receive exclusive updates, interviews, discounts, and more. Join Us!	26
source Add To The Conversation Using Facebook Comments	26
...	...
A new study that explores the effects of vouchers on Catholic parishes and their schools finds that while vouchers may help keep a church financially afloat, the expansion of government funding leads to a decline in private donations from parishioners. [The study, a working paper by the National Bureau of Economic Research (NBER) is titled "Beyond the Classroom: The Implications of School Vouchers for Church Finances. " Notre Dame University economics associate professor Daniel Hungerman and colleagues Kevin Rinz, and Jay Frymark analyzed financial data from Catholic parishes in Milwaukee, Wisconsin, using information from both the parishes and their schools. "We show that vouchers are now a dominant source of funding for many churches," the researchers write. "Parishes in our sample running schools get more revenue from vouchers than from worshipers. We also find that voucher expansion prevents church closures and mergers. " The authors continue: Despite these results, we fail to find evidence that vouchers promote religious behavior: voucher expansion causes significant declines in church donations and church spending on religious purposes. The meteoric growth of vouchers appears to offer financial stability for congregations while at the same time diminishing their religious activities. While voucher programs for Catholic schools have already been on the rise in nearly half of all states in America, the study is released as the topic of school choice and school vouchers has risen to national prominence with the election of President Donald Trump. On Friday, Trump and U. S. Education Secretary Betsy DeVos visited St. Andrew Catholic School in Pine Hills, Florida, to highlight the administration's plan to offer children from poorly performing public schools other education opportunities. In his address to a joint session of Congress Tuesday, Trump said, "Education is the civil rights issue of our time. " "I am calling upon Members of both parties to pass an education bill that funds school choice for disadvantaged youth, including millions of and Latino children," he added. Catholic schools are the most common type of religious private school in the United States. Nevertheless, these schools and the parishes that support them have faced a significant economic downturn. In 2011, the National Catholic Educational Association (NCEA) reported that, between 2000 and 2010, the number of Catholic schools dropped from 8, 146 to 6, 980 — a loss of 117 schools per year. Additionally, Catholic elementary and secondary school enrollment fell 22 percent during that same period, from 2, 647, 301 to 2, 065, 872. The NBER researchers observe: Brinig and Garnett (2014) note several reasons for dwindling enrollment, including competition from charter schools, struggles to appeal to populations (e. g. Hispanics) and increasing tuition costs. They summarize the situation of Catholic schools thusly: "The Catholic school financial model — which depended upon the generosity of parishioners in pews that are now empty and the free labor of nuns who are now retired — cannot be sustained. " Catholic schools, the authors continue, have turned to diverse populations, including students from families, to remain open. The government funding from vouchers has had an impact on both the schools and their parishes, they add. The researchers assert: We find that expansion in voucher policy is, unsurprisingly, associated with increases in voucher revenues for parishes with schools. We also find that voucher expansion prevents parish closures and mergers. … We further see how voucher expansion impacts parishes' finances. Interestingly, and the above results notwithstanding, we find no evidence that vouchers subsidize parish religious activity beyond the operation of religious schools. In fact, our estimates repeatedly suggest	1

```
df['title'].value_counts()
```

	count
title	
No Title Available	518
a news release	240
an interview	229
a press release	223
a speech	214
...	...
At Euro 2016 Final, Portugal Loses Ronaldo but Defeats France - The New York Times	1
a flier mailed to constituents	1
EPIC! TUCKER CARLSON Demolishes NYC Councilman Over Sanctuary Cities [Video]	1
'The Greatest Loss and Despair': Thais Mourn Their King - The New York Times	1
Progressive Nomad: Bernie Sanders Pens CRUCIAL Letter To The President (IMAGES)	1

24149 rows × 1 columns

role for the militia in the Mosul offensive. Tal Afar, north of Mosul city, is a predominantly Turkmen town that was captured by ISIS two years ago when the group captured

```
df.dtypes
```

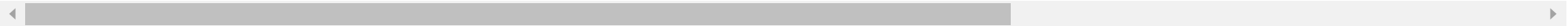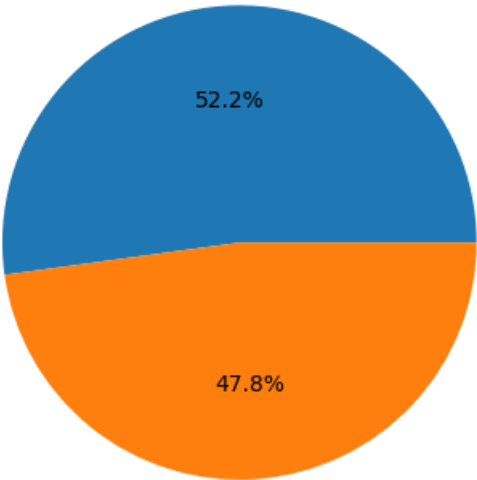|	0
title	object
author	object
text	object
label	int64

```
df_new=df.copy()
```

```
df.reset_index(drop=True,inplace=True)
```

> and cried out to all good men. "Join me!". And the people joined. From village to village across all the lands, they joined. \nAnd on a dark windy night, the people were all

```
import matplotlib.pyplot as plt
```

> was coming. They prayed for his safety. \nThey huddled together in fear as they felt the power of his chariot coming from the sky. \nHis chariot stopped in front of the

```
plt.pie(df["label"].value_counts(),autopct='%1.1f%%')
plt.show()
```


Remove Punctuation

```
import string
punc=string.punctuation
```

Remove Stopwords

```
import  nltk
nltk.download("stopwords")
from nltk.corpus import stopwords
stop_words=stopwords.words("english")
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
```

Tokenization

```
from nltk.tokenize import word_tokenize
nltk.download("punkt_tab")
```

```
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt_tab.zip.
True
```

Lemmatization

```
from nltk.stem import WordNetLemmatizer
nltk.download("wordnet")
lemmatizer=WordNetLemmatizer()
```

```
[nltk_data] Downloading package wordnet to /root/nltk_data...
```

```
def preprocess(sentence):
  sentence1=sentence.lower()
  sentence2=[i for i in sentence1 if i not in punc]
  sentence3=''.join(sentence2)
  tokens=word_tokenize(sentence3)
  sentence4=[i for i in tokens if i not in stop_words]
  sentence6=[lemmatizer.lemmatize(i,'v') for i in sentence4]
  return " ".join(sentence6)
```

```
df
```

	title	author	text	label
0	a news release	federation-american-immigration-reform	Unemployment has been on the rise throughout W...	1
1	Black Turnout Soft in Early Voting, Boding Ill...	Henry Wolff	Black Turnout Soft in Early Voting, Boding Ill...	1
2	a television interview	chris-abele	Says Milwaukee County buses are no less safe n...	1
3	10 Things I Learned From Being My Own General ...	Luke Stranahan	One of the foundations of living a good life i...	1
4	a statement responding to Gov. Rick Scott's St...	rod-smith	Says Rick Scotts proposed budget would lay off...	0
...	...	...	...	...
30823	a TV ad.	bill-richardson	Bill Richardson passed the first law in the na...	0
30824	Khodorkovsky: Putin is not going to cozy up to...	Unknown Author	Print version Font Size The number of predicti...	1
30825	The Only Coup Is a Stolen Election	Dave Hodges	\n\nWars and rumor of wars. There is no coup. ...	1
30826	Doomed Jet Carrying Brazilian Team Reportedly ...	Nicholas Casey and Ewan MacKenna	The pilot of a plane that crashed late Monday ...	0
30827	Burger King 'O.K. Google' Ad Doesn't Seem O.K....	Sapna Maheshwari	Google appeared to stymie a marketing stunt on...	0

30828 rows × 4 columns

```
df["News"]=df["title"]+" "+df['text']
df['News'] = df['News'].apply(preprocess)
```

```
df['News']
```

	News
0	news release unemployment rise throughout wisc...
1	black turnout soft early vote bode ill hillary...
2	television interview say milwaukee county bus ...
3	10 things learn general contractor one foundat...
4	statement respond gov rick scotts state state ...
...	...
30823	tv ad bill richardson pass first law nation gi...
30824	khodorkovsky putin go cozy washington print ve...
30825	coup steal election war rumor war coup watch m...
30826	doom jet carry brazilian team reportedly run f...
30827	burger king ' ok google ' ad ' seem ok google ...

30828 rows × 1 columns

```
y=df['label']
```

```
from sklearn.feature_extraction.text import CountVectorizer
vectorizer=CountVectorizer()
sentence=vectorizer.fit_transform(df['News'])
```

```
sentence.shape
```

```
(30828, 192409)
```

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(sentence,y,test_size=0.3,random_state=42)
```

```
from sklearn.naive_bayes import MultinomialNB
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier,GradientBoostingClassifier
from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
```

```
models=[MultinomialNB(),SVC(),RandomForestClassifier(),GradientBoostingClassifier()]
for i in models:
  model=i
  model.fit(x_train,y_train)
  y_pred=model.predict(x_test)
  print(model)
  print()
  print(f"Accuracy score: {round((accuracy_score(y_test,y_pred))*100)}%")
  print()
  print(confusion_matrix(y_test,y_pred))
  print(classification_report(y_test,y_pred))
```

```
Accuracy score: 79%

[[4366  476]
 [1457 2950]]
              precision    recall  f1-score   support

           0       0.75      0.90      0.82      4842
           1       0.86      0.67      0.75      4407

    accuracy                           0.79      9249
```

```
SVC()

Accuracy score: 80%

[[3414 1428]
 [ 413 3994]]
              precision    recall  f1-score   support

           0       0.89      0.71      0.79      4842
           1       0.74      0.91      0.81      4407

    accuracy                           0.80      9249
   macro avg       0.81      0.81      0.80      9249
weighted avg       0.82      0.80      0.80      9249

RandomForestClassifier()

Accuracy score: 83%

[[4307  535]
 [1045 3362]]
              precision    recall  f1-score   support

           0       0.80      0.89      0.85      4842
           1       0.86      0.76      0.81      4407

    accuracy                           0.83      9249
   macro avg       0.83      0.83      0.83      9249
weighted avg       0.83      0.83      0.83      9249

GradientBoostingClassifier()

Accuracy score: 82%

[[3596 1246]
 [ 455 3952]]
              precision    recall  f1-score   support

           0       0.89      0.74      0.81      4842
           1       0.76      0.90      0.82      4407

    accuracy                           0.82      9249
   macro avg       0.82      0.82      0.82      9249
weighted avg       0.83      0.82      0.82      9249
```

```
from sklearn.model_selection import cross_val_score
cross_val_score(RandomForestClassifier(),sentence,y,cv=5).mean()
```

```
0.8310628491697267
```

```
from sklearn.model_selection import GridSearchCV,RandomizedSearchCV
parameters = {'n_estimators': [50, 100, 150, 200],'max_depth': [None, 10, 20, 30],'min_samples_split': [2, 5, 10],'min_samples_leaf': [1, 2, 4]}
model = RandomizedSearchCV(RandomForestClassifier(),parameters,cv=5,random_state=42,n_iter=10)
model.fit(sentence, y)
```

```
        ▸          RandomizedSearchCV               ⓘ  ?

        ▸          best_estimator_:
                   RandomForestClassifier

          ▸  RandomForestClassifier      ?
```

```
model.best_params_
```

```
{'n_estimators': 200,
 'min_samples_split': 5,
 'min_samples_leaf': 2,
 'max_depth': None}
```

```
model.best_score_
```

```
0.8325875224493264
```

```
print(model)
```

```
RandomizedSearchCV(cv=5, estimator=RandomForestClassifier(),
                   param_distributions={'max_depth': [None, 10, 20, 30],
                                        'min_samples_leaf': [1, 2, 4],
                                        'min_samples_split': [2, 5, 10],
                                        'n_estimators': [50, 100, 150, 200]},
                   random_state=42)
```

```
model=RandomForestClassifier()#"n_estimators":150,"min_samples_split":5,"min_samples_leaf":2)
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
print(model)
print()
print(f"Accuracy score: {round((accuracy_score(y_test,y_pred))*100)}%")
print()
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
```

```
RandomForestClassifier()

Accuracy score: 83%

[[4271  571]
 [ 986 3421]]
```

	precision	recall	f1-score	support
0	0.81	0.88	0.85	4842
1	0.86	0.78	0.81	4407
accuracy			0.83	9249
macro avg	0.83	0.83	0.83	9249
weighted avg	0.83	0.83	0.83	9249

```python
def predict(sentence):
  sentence1=preprocess(sentence)
  vector=vectorizer.transform([sentence1])
  pred=model.predict(vector)
  if pred==0:
    print("Real News")
  else :
    print("Fake News")
```

```python
sentence="India successfully test-fired two surface-to-air missiles within 48 hours, enhancing its air defense capabilities. The tests demonstrate India's commitment t
```

```python
predict(sentence)
```

    Real News

```python
predict("Scientists have discovered a new drug that can cure all forms of cancer, and it's available for free to the public starting tomorrow.")
```

    Fake News