Name of Project : Fake_news_detection

Member: Hena Razbaan , Yalda Akbari

Department: IS

Semester: 5

The Fake news detection problem using four machine learning classification algorithms. This project related apply linear regression Decision Tree classification, Gradient boost classification and random forest classification model.

What is manual testing?

The purpose of manual testing is to catch bugs and feature issues before a software application goes live. When manually testing, the tester validates the key features of a software application. Analysts execute test cases and develop summary error reports without specialized automation tools.

What is inplace?

Inplace is a parameter accepted by a number of panda's methods which affects the behavior of how the method runs. Some examples of where you might commonly see this keyword (but hopefully not implemented in your own code) are the methods; (fillna), (Replace), (Rename), the list goes on

What is pd.concat?

Advertisements. Pandas provides various facilities for easily combining together Series, DataFrame, and Panel objects. Pd.concat(objs,axis=0,join='outer',join_axes=None, ignore_index=False) objs – This is a sequence or mapping of Series, DataFrame, or Panel objects

What is Marge ?

Data merging is the process of combining two or more data sets into a single data set. Most often, this process is necessary when you have raw data stored in multiple files, worksheets, or data tables, that you want to analyze all in one go.

Here:

Removing last 10 rows from both the dataset , for manual testing .

A value is trying to be set on a copy of a slice from a dataframe .

Try Using .loc [row_index , col_indexer]=value instead

df_fake_manual_testing["class"]=0

A value is trying to be set on a copy of a slice from a Dataframe .

Try using . loc [row_index , col_indexer]=value instead

df_true_manual_testing["class"]=1

	title	text	subject	date	class
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017	0
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017	0
7	Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017	0
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017	0
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we d get a br...	News	December 21, 2017	0

"title', "subject" and "date" columns is not required for detecting the fake news , so I am going to drop the columns.

Drop: remove

Frac:

Int value, Number of random rows to generate. Frac: Float value, Returns (float value * length of data frame values). Frac cannot be used with n. replace: Boolean value, return sample with replacement if True. Random_state: int value or numpy

Sample(frac =" ") method returns a list with a randomly selection of a specified number of items from a sequnce.

Return a random sample of items from an axis of object

```
In [15]: df = df.sample(frac = 1)
         df.head(10)
```

Out[15]:

	text	class
2789	SAN FRANCISCO (Reuters) - The laws that prohib...	1
18363	French President Emmanuel Macron has prided hi...	0
6287	FRANKFURT/BERLIN (Reuters) - U.S President-ele...	1
1655	(Reuters) - Wal-Mart Stores Inc, Target Corp a...	1
19696	Just look away. The Democrats don t have any p...	0
22357	Charles Hugh Smith Of Two MindsSo when the pr...	0
2221	WASHINGTON (Reuters) - The head of the U.S. Se...	1
58	WASHINGTON (Reuters) - The top Democrat on the...	1
15153	TOKYO (Reuters) - Japan will build four coast ...	1
15070	The first shot against PC tyranny in Virginia ...	0

```
In [16]: df.isnull().sum()
```

Out[16]: text   0

Apply : a function along an axis of the DataFrame

The isnull() method returns a DataFrame object where all the values are replaced with a Boolean value True for NULL values, and otherwise False

Here:

Creating a function to convert the text lowercase , remove the extra space , special chr , url and links.

Muc...   br...   2017

```
In [14]: df = df_marge.drop(["title", "subject","date"], axis = 1)
         df.head(10)
```

Out[14]:

	text	class
0	Donald Trump just couldn t wish all Americans ...	0
1	House Intelligence Committee Chairman Devin Nu...	0
2	On Friday, it was revealed that former Milwauk...	0
3	On Christmas day, Donald Trump announced that ...	0
4	Pope Francis used his annual Christmas Day mes...	0
5	The number of cases of cops brutalizing and ki...	0
6	Donald Trump spent a good portion of his day a...	0
7	In the wake of yet another court decision that...	0
8	Many people have raised the alarm regarding th...	0
9	Just when you might have thought we d get a br...	0

```
In [15]: df = df.sample(frac = 1)
         df.head(10)
```

In [19]: `df.head(10)`

Out[19]:

	text	class
2789	san francisco reuters the laws that prohib...	1
18363	french president emmanuel macron has prided hi...	0
6287	frankfurt berlin reuters u s president ele...	1
1655	reuters wal mart stores inc target corp a...	1
19696	just look away the democrats don t have any p...	0
22357	charles hugh smith of two mindsso when the pr...	0
2221	washington reuters the head of the u s se...	1
58	washington reuters the top democrat on the...	1
15153	tokyo reuters japan will build four coast ...	1
15070	the first shot against pc tyranny in virginia ...	0

In [20]:
```
x = df["text"]
y = df["class"]
```

Logistic Regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.).

logistic Regression

In [24]: `from sklearn.linear_model import LogisticRegression`

In [25]:
```
LR = LogisticRegression()
LR.fit(xv_train,y_train)
```

Out[25]: LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                   intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
                   penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
                   verbose=0, warm_start=False)

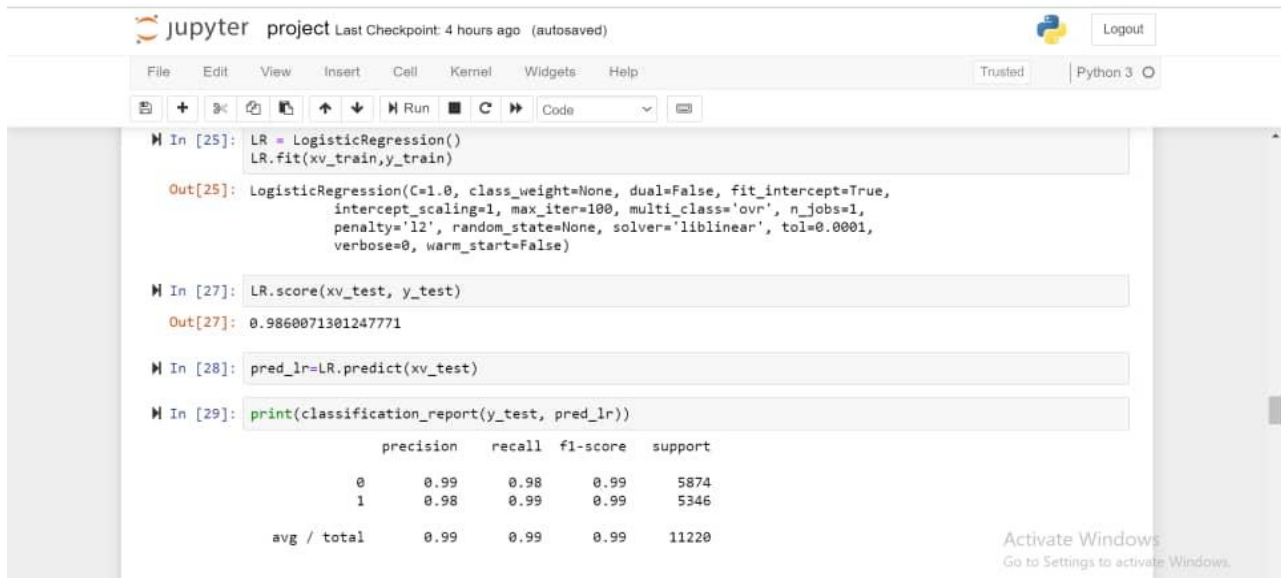In [27]: `LR.score(xv_test, y_test)`

Out[27]: 0.9860071301247771

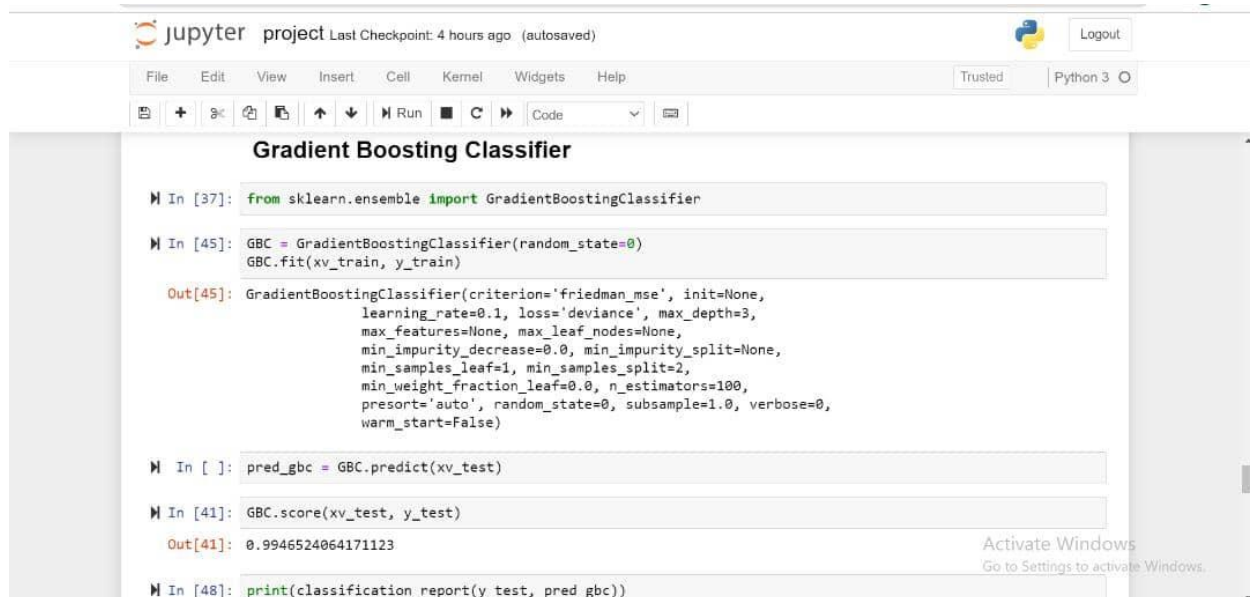In [28]: `pred_lr=LR.predict(xv_test)`

In [29]: `print(classification_report(y_test, pred_lr))`

```
              precision    recall  f1-score   support
```

```
In [25]: LR = LogisticRegression()
         LR.fit(xv_train,y_train)

Out[25]: LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                   intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
                   penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
                   verbose=0, warm_start=False)
```

```
In [27]: LR.score(xv_test, y_test)

Out[27]: 0.9860071301247771
```

```
In [28]: pred_lr=LR.predict(xv_test)
```

```
In [29]: print(classification_report(y_test, pred_lr))

                 precision    recall  f1-score   support

             0       0.99      0.98      0.99      5874
             1       0.98      0.99      0.99      5346

    avg / total       0.99      0.99      0.99     11220
```

Decision tree is a type of supervised learning algorithm that can be used for both regression and classification problems. The algorithm uses training data to create rules that can be represented by a tree structure. Like any other tree representation, it has a root node, internal nodes, and leaf nodes.

Gradient boosting classifiers are a group of machine learning algorithms that combine many weak learning models together to create a strong predictive model. Decision trees are usually used when doing gradient boosting.

Gradient Boosting Classifier

```
In [37]:  from sklearn.ensemble import GradientBoostingClassifier
```

```
In [45]:  GBC = GradientBoostingClassifier(random_state=0)
          GBC.fit(xv_train, y_train)
```
```
Out[45]:  GradientBoostingClassifier(criterion='friedman_mse', init=None,
                     learning_rate=0.1, loss='deviance', max_depth=3,
                     max_features=None, max_leaf_nodes=None,
                     min_impurity_decrease=0.0, min_impurity_split=None,
                     min_samples_leaf=1, min_samples_split=2,
                     min_weight_fraction_leaf=0.0, n_estimators=100,
                     presort='auto', random_state=0, subsample=1.0, verbose=0,
                     warm_start=False)
```
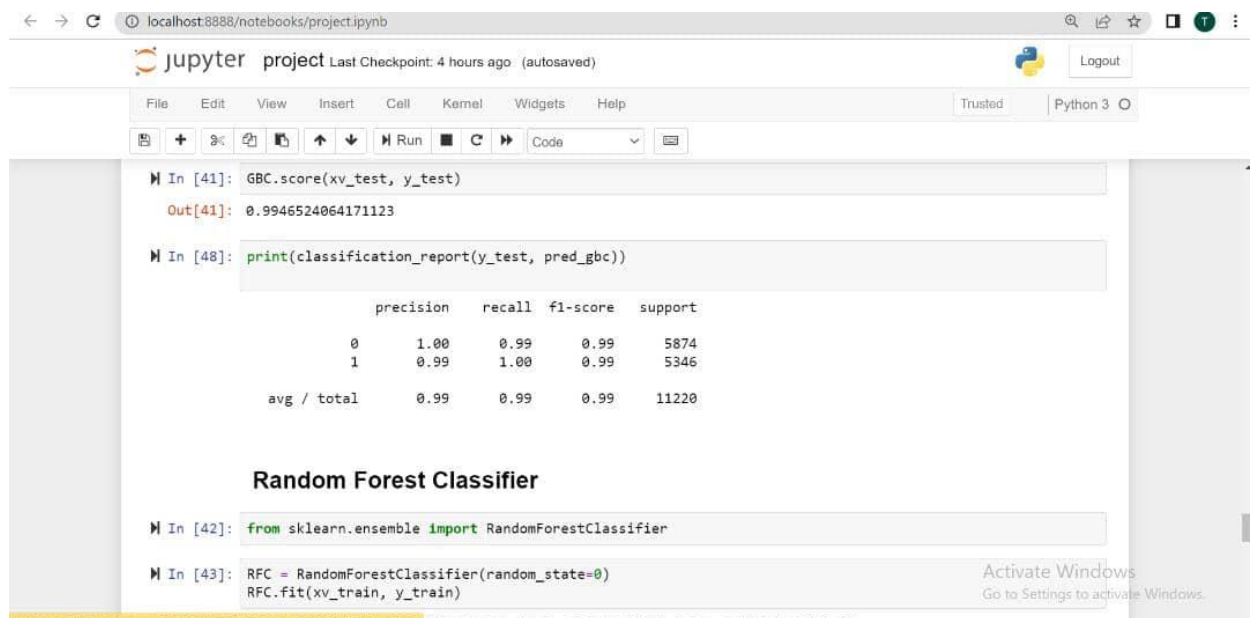```
In [ ]:   pred_gbc = GBC.predict(xv_test)
```
```
In [41]: GBC.score(xv_test, y_test)
```
```
Out[41]: 0.9946524064171123
```
```
In [48]: print(classification_report(y_test, pred_gbc))
```

localhost:8888/notebooks/project.ipynb

```
In [41]: GBC.score(xv_test, y_test)
```
```
Out[41]: 0.9946524064171123
```
```
In [48]: print(classification_report(y_test, pred_gbc))
```

```
                precision    recall  f1-score   support

            0        1.00      0.99      0.99      5874
            1        0.99      1.00      0.99      5346

  avg / total        0.99      0.99      0.99     11220
```

Random Forest Classifier

```
In [42]:  from sklearn.ensemble import RandomForestClassifier
```

```
In [43]:  RFC = RandomForestClassifier(random_state=0)
          RFC.fit(xv_train, y_train)
```

The random forest is a classification algorithm consisting of many decisions trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree.

Random Forest Classifier

```
In [42]:  from sklearn.ensemble import RandomForestClassifier
```

```
In [43]:  RFC = RandomForestClassifier(random_state=0)
          RFC.fit(xv_train, y_train)
```

```
Out[43]:  RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
                      max_depth=None, max_features='auto', max_leaf_nodes=None,
                      min_impurity_decrease=0.0, min_impurity_split=None,
                      min_samples_leaf=1, min_samples_split=2,
                      min_weight_fraction_leaf=0.0, n_estimators=10, n_jobs=1,
                      oob_score=False, random_state=0, verbose=0, warm_start=False)
```

```
In [46]:  pred_rfc = RFC.predict(xv_test)
```

```
In [47]:  RFC.score(xv_test, y_test)
```

```
Out[47]:  0.9679144385026738
```

```
In [49]:  print(classification_report(y_test, pred_rfc))
```

localhost:8888/tree?token=b53d5632f4ddc389643bcaa020e965257141af63d... ion   recall f1-score   support

---

```
                     min_weight_fraction_leaf=0.0, n_estimators=10, n_jobs=1,
                     oob_score=False, random_state=0, verbose=0, warm_start=False)
```

```
In [46]:  pred_rfc = RFC.predict(xv_test)
```
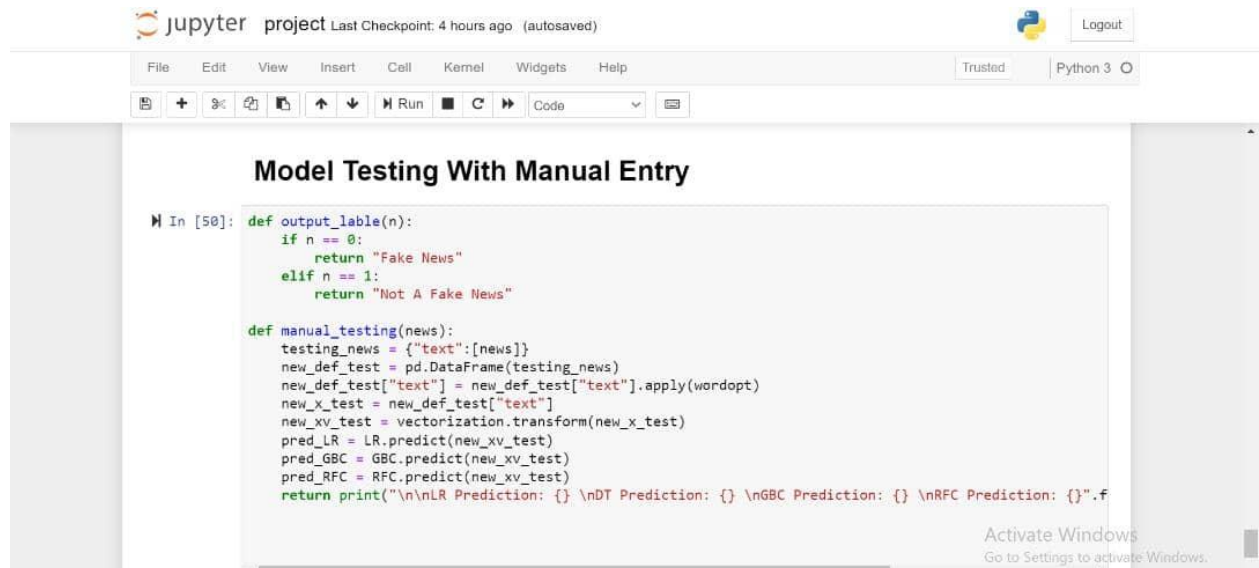
```
In [47]:  RFC.score(xv_test, y_test)
```

```
Out[47]:  0.9679144385026738
```

```
In [49]:  print(classification_report(y_test, pred_rfc))

                    precision    recall  f1-score   support

                0       0.95      0.99      0.97      5874
                1       0.99      0.95      0.97      5346

        avg / total     0.97      0.97      0.97     11220
```
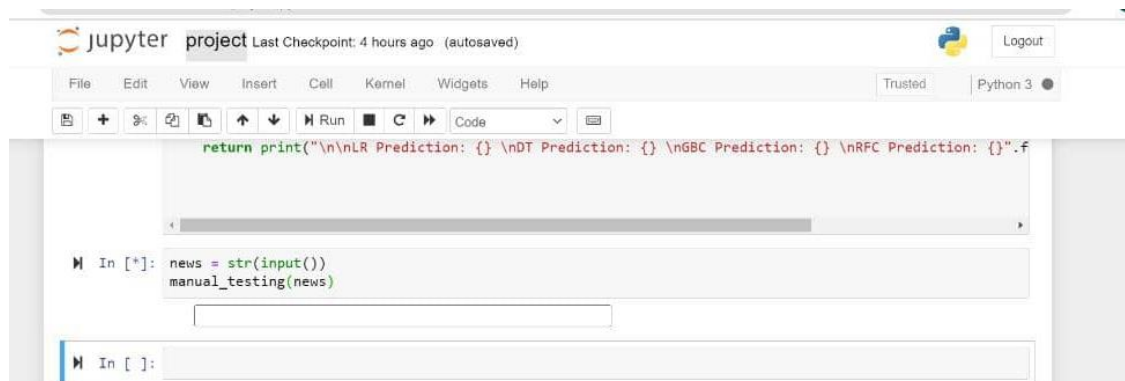
When we import the title on input, & we are getting the title from manual-testing, so the output will going to declare us which the news are df_true or df_false.

Model Testing With Manual Entry

```python
In [50]: def output_lable(n):
             if n == 0:
                 return "Fake News"
             elif n == 1:
                 return "Not A Fake News"

         def manual_testing(news):
             testing_news = {"text":[news]}
             new_def_test = pd.DataFrame(testing_news)
             new_def_test["text"] = new_def_test["text"].apply(wordopt)
             new_x_test = new_def_test["text"]
             new_xv_test = vectorization.transform(new_x_test)
             pred_LR = LR.predict(new_xv_test)
             pred_GBC = GBC.predict(new_xv_test)
             pred_RFC = RFC.predict(new_xv_test)
             return print("\n\nLR Prediction: {} \nDT Prediction: {} \nGBC Prediction: {} \nRFC Prediction: {}".f
```
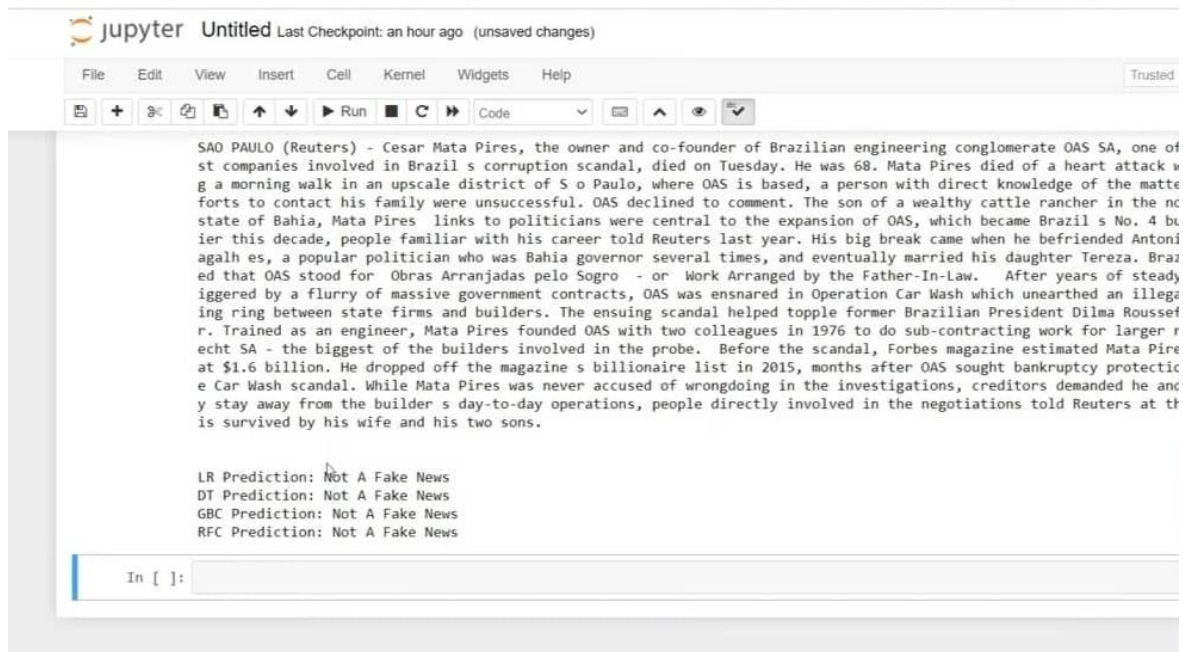
```python
         return print("\n\nLR Prediction: {} \nDT Prediction: {} \nGBC Prediction: {} \nRFC Prediction: {}".f
```

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

```
SAO PAULO (Reuters) - Cesar Mata Pires, the owner and co-founder of Brazilian engineering conglomerate OAS SA, one of
st companies involved in Brazil s corruption scandal, died on Tuesday. He was 68. Mata Pires died of a heart attack w
g a morning walk in an upscale district of S o Paulo, where OAS is based, a person with direct knowledge of the matte
forts to contact his family were unsuccessful. OAS declined to comment. The son of a wealthy cattle rancher in the no
state of Bahia, Mata Pires  links to politicians were central to the expansion of OAS, which became Brazil s No. 4 bu
ier this decade, people familiar with his career told Reuters last year. His big break came when he befriended Antoni
agalh es, a popular politician who was Bahia governor several times, and eventually married his daughter Tereza. Braz
ed that OAS stood for  Obras Arranjadas pelo Sogro  - or  Work Arranged by the Father-In-Law.   After years of steady
iggered by a flurry of massive government contracts, OAS was ensnared in Operation Car Wash which unearthed an illega
ing ring between state firms and builders. The ensuing scandal helped topple former Brazilian President Dilma Roussef
r. Trained as an engineer, Mata Pires founded OAS with two colleagues in 1976 to do sub-contracting work for larger r
echt SA - the biggest of the builders involved in the probe.  Before the scandal, Forbes magazine estimated Mata Pire
at $1.6 billion. He dropped off the magazine s billionaire list in 2015, months after OAS sought bankruptcy protectio
e Car Wash scandal. While Mata Pires was never accused of wrongdoing in the investigations, creditors demanded he and
y stay away from the builder s day-to-day operations, people directly involved in the negotiations told Reuters at th
is survived by his wife and his two sons.


LR Prediction: Not A Fake News
DT Prediction: Not A Fake News
GBC Prediction: Not A Fake News
RFC Prediction: Not A Fake News
```