T5 BOOTSCAMP DATA SCIENCE PROJECT

FAKE NEWS DETECTION

Ghaida Faisal Alharbi | Project 1 | T5o06

Project Goal

Classify whether news is true or fake using machine learning models

DATA SOURCES

The datasets comes from Kaggle, it is provided in .csv format.

- **True.csv**

	title	text	subject	date
0	"As U.S. budget fight looms, Republicans flip their fiscal script"	"WASHINGTON (Reuters) -The head of a conservative Republican faction in ..."	politicsNews	December 31,2017

- **Fake.csv**

	title	text	subject	date
0	"DonaldTrump Sends OutEmbarrassing New Year's..."	"Donald Trump just couldn t wish all Americans a Happy New Year and leave it at that..."	News	December 31,2017

Exploratory Data Analysis

IMPORTING LIBRARIES

Pandas, Matplotlib, and seaborn..etc

READ DATASETS

read True.csv and Fake.csv file , and take a look using head(), info(), and describe for each file.

COMBINE DATASETS

first, I create new column for each Datasets called " Label" , then I will combine them using concat() method.

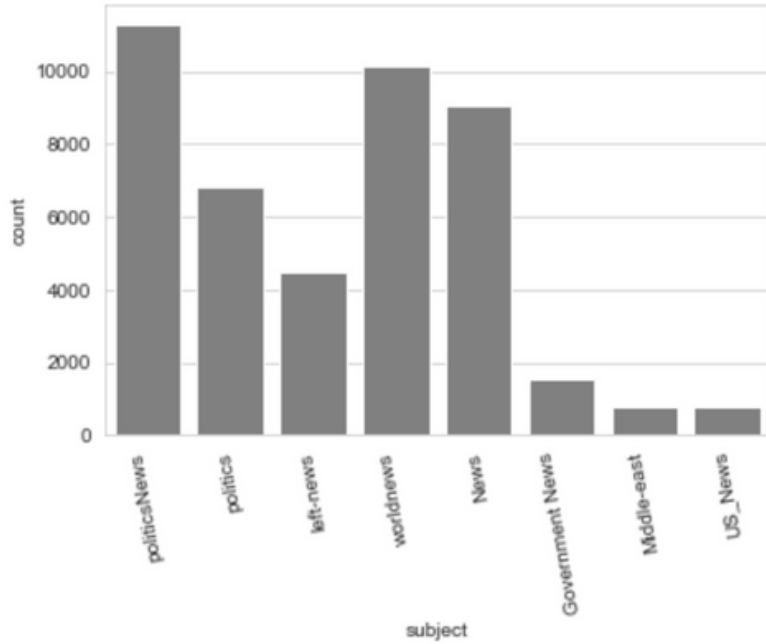
CLEANING DATAFRAME

Remove duplicated recordes , remove columns, and process text.

Visualization the data

COUNT SUBJECT

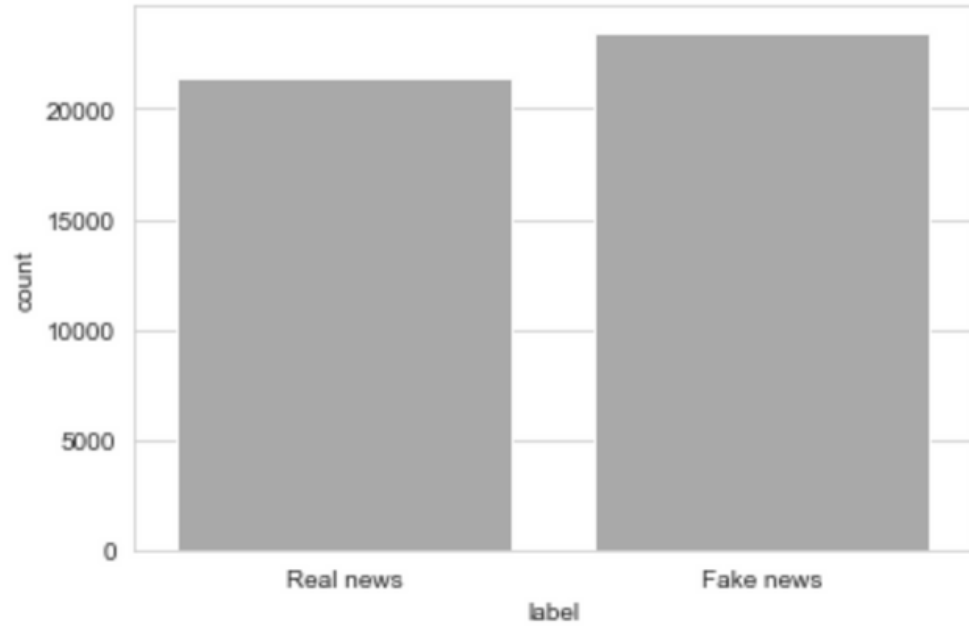
```
Out[58]: (array([0, 1, 2, 3, 4, 5, 6, 7]),
          [Text(0, 0, 'politicsNews'),
           Text(1, 0, 'politics'),
           Text(2, 0, 'left-news'),
           Text(3, 0, 'worldnews'),
           Text(4, 0, 'News'),
           Text(5, 0, 'Government News'),
           Text(6, 0, 'Middle-east'),
           Text(7, 0, 'US_News')])
```



Visualization the data

THE DISTRIBUTION OF REAL AND FAKE NEWS

```
[59]: <AxesSubplot:xlabel='label', ylabel='count'>
```



LOGISTIC REGRESSION

RANDOM FOREST

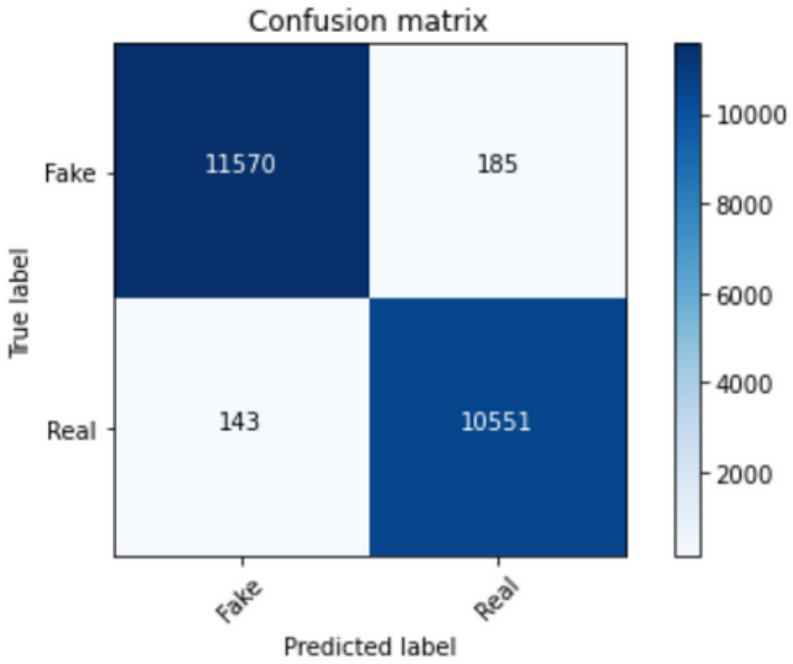
MODELING

NAIVE-BAYAS

K-NEAREST NIEGHBORS (KNN)

CODE

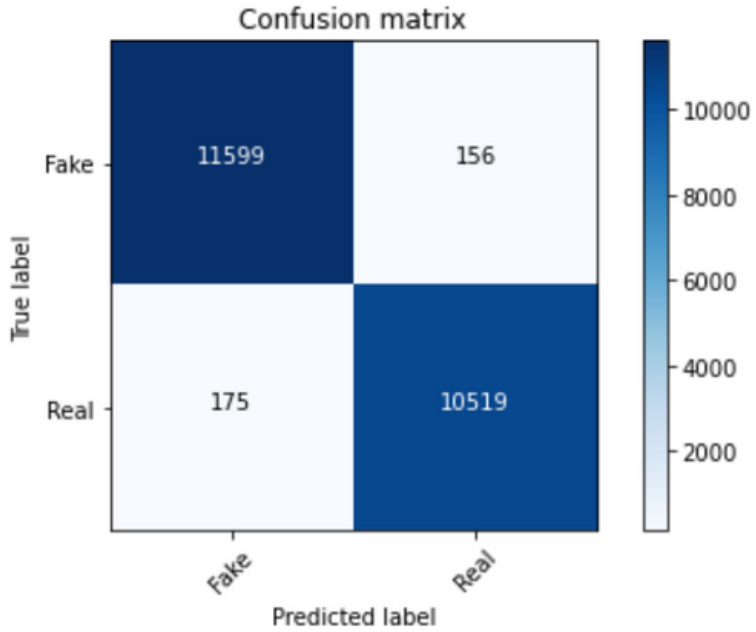
	precision	recall	f1-score	support
Fake news	0.99	0.98	0.99	11755
Real news	0.98	0.99	0.98	10694
accuracy			0.99	22449
macro avg	0.99	0.99	0.99	22449
weighted avg	0.99	0.99	0.99	22449



Logistic Regression

CODE

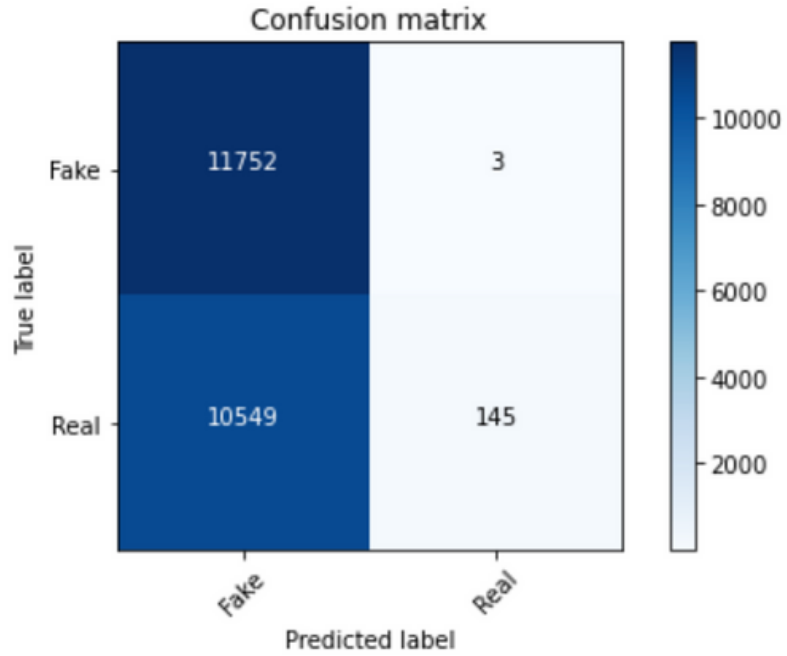
	precision	recall	f1-score	support
Fake news	0.99	0.99	0.99	11755
Real news	0.99	0.98	0.98	10694
accuracy			0.99	22449
macro avg	0.99	0.99	0.99	22449
weighted avg	0.99	0.99	0.99	22449



RFC

CODE

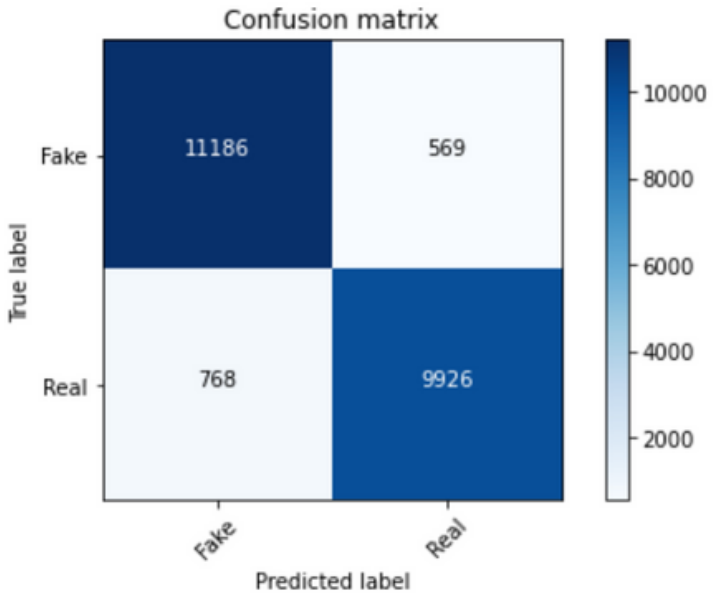
	precision	recall	f1-score	support
Fake news	0.53	1.00	0.69	11755
Real news	0.98	0.01	0.03	10694
accuracy			0.53	22449
macro avg	0.75	0.51	0.36	22449
weighted avg	0.74	0.53	0.37	22449



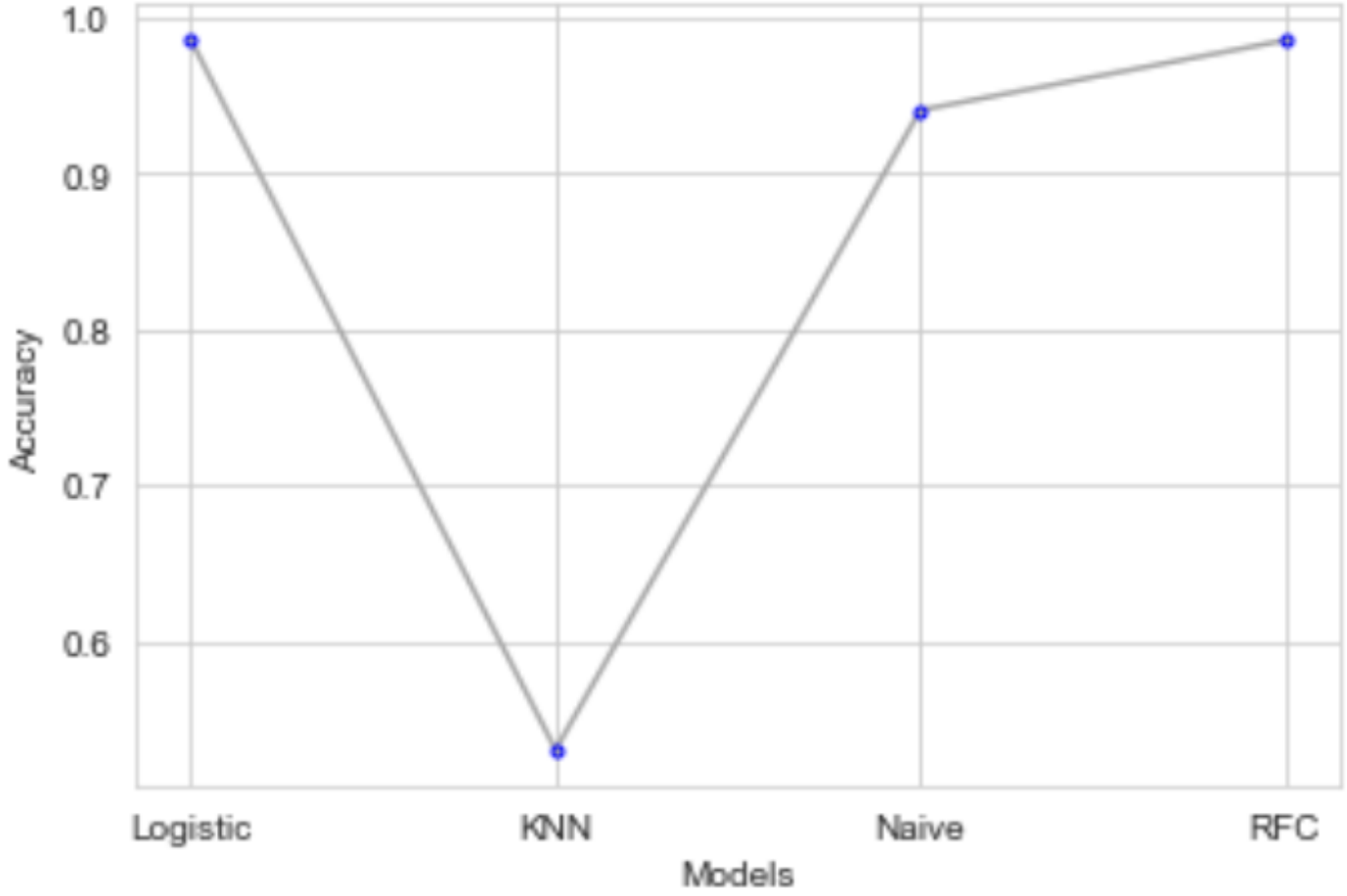
KNN

CODE

	precision	recall	f1-score	support
Fake news	0.94	0.95	0.94	11755
Real news	0.95	0.93	0.94	10694
accuracy			0.94	22449
macro avg	0.94	0.94	0.94	22449
weighted avg	0.94	0.94	0.94	22449



Naive-bayas



COMPARING

comparing between models based on accuracy score

Any Questions ?

Thanks!