FAKE INFORMATION DETECTION

A capstone project report submitted by

AMRUTHA D S

in partial fulfillment for the award of the program

DATA SCIENCE BOOTCAMP

in

UPGRAD KNOWLEDGEHUT

Email ID: amruthads2002@gmail.com

CONTENTS

ABSTRACT

TITLE: FAKE INFORMATION DETECTION

This project is solely based on the purpose of creating a fake news detector for a given data set. This is a project associated with data analysis that was created using python programming language along with which machine learning classification algorithms namely Linear aggression, Decision tree classification, Gradient boost classification and Random Forest classification model were used.

We live in the era where people blindly believe rumors and do not think twice before turning it into a gossip session without thoroughly checking the facts. Any kind of news no matter what gets spread quickly irrespective of time and distance.

Fake news, alternative facts are associated to each other since the time news was transmitted using newspapers or radio. There have been several hoax stories where citizens, governments as well all other social elements are all affected by these kinds of fake stories. Several social media organizations have been subjected to controversies by the media houses for targeting the audiences and showing them posts to their support.

This project mainly focuses on detecting fake news with the help of various python libraries in association with counting feature such as Tfidf Vectorizer. The system will be taking input from the user and then compare them with an existing data-set. I have compared various algorithms to find out the best working model that will fit our project and give a proper prediction for fake news.

The main objective is to detect the fake news, which is a classic text classification problem with a straightforward proposition. It is needed to build a model that can differentiate between “Real” news and “Fake” news thus helping in getting the facts out right.

KEY-WORDS: Linear aggression, Decision tree classification, Gradient boost classification, Random Forest classification.

CHAPTER 1

INTRODUCTION

1.1 Introduction

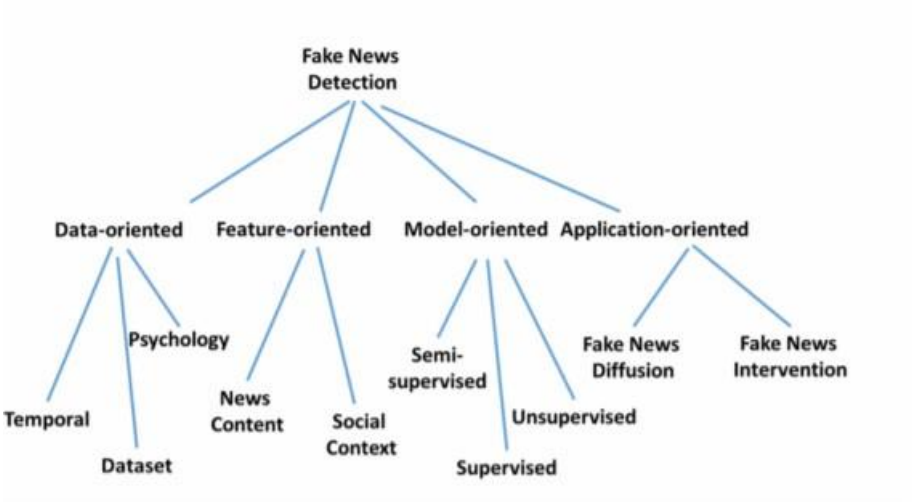

Fig.1.1 Fake news detection model

Fake news detection has gained a great deal of interest and popularity among the massesespecially from researchers around the world. There are numerous studies that have been conducted on the impact of fake news and how humans react to them. Fake news can be any content that is not truthful and generated to convince its readers to believe in something that is not true. In today's times there are various social media messaging and share applications that give users the power to share a piece of information with millions of people at the click of the button. The real problem is when people start to accept that rather than any of the news being "fake" theirs might have a new perspective on this. The problem begins where the masses begin to believe the fake news without checking its authenticity. There are very few tools or websites that tell the public about the news and its authenticity.

These days' fake news is creating different issues from sarcastic articles to a fabricated news and plan government propaganda in some outlets. Fake news and lack of trust in the media are growing problems with huge ramifications in our society. Obviously, a purposely misleading story is "fake news" but lately blathering social media's discourse is changing its definition. Some of them now use the term to dismiss the facts counter to their preferred viewpoints.

Facebook, Instagram, Twitter and many other social media platforms has been at the epicenter of much critique following media attention. They have already implemented a feature to flag fake news on the site when a user comes across such articles. They have also said publicly that they are working on to distinguish these articles in an automated way. Certainly, it is not an easy

task. A given algorithm must be politically unbiased since fake news exists on both ends of the spectrum and also gives equal balance to legitimate news sources on either end of the spectrum. In addition, the question of legitimacy is a difficult one. However, in order to solve this problem, it is necessary to have an understanding on what Fake News is. Later, it is needed to look in to how the techniques in the field of machine learning help us to detect fake news.

1.2 Objectives

The main objective of this project is to detect the fake news, which is a classic text classification problem with a straightforward approach. It is needed to build a model that can differentiate between “Real” news and “Fake” news. A data set containing both the kind of news will be provided which will in turn be used to classify them into fake or real news using Machine Learning algorithms with the help of Python programming language.

1.3 Motivation

During these tough times as, our World is going through a pandemic and I am sure all of us would have come across certain news related to virus that was interpreted in a manner which seemed to be true only to realize later with certain clarifications that it was false or fake. Andalso, life as such is an enormous journey wherein all of us will encounter certain uncertainty inthe way and blindly follow the path without even analyzing it and not knowing it to be true or false. So these kind of situations helped me build an idea of creating this project. Through thorough research and analysis, I could build this project with the technical concepts which will help the user to differentiate between fake news and real news.

1.4 Overview of the project

A dataset known as news.csv was acquired which in turn had two kinds of news within itself that was classified as True news and Fake news .Using jupyterlab the python code has been executed in which the Machine learning algorithms such as Linear aggression, Decision tree classification, Gradient boost classification and Random forest classification were used .TfidfVectorizer was used in the program to detect the word frequency scores and it will tokenize documents, learn the vocabulary and inverse document frequency weightings and allows the user to encode new documents.

CHAPTER 2

ANALYSIS AND DESIGN

2.1 Functional Requirements

This project can provide the user with a basic degree of functionality and the user experience provided would be smooth enough to differentiate between fake and real news. Moreover, the algorithms used for this program are functional enough to verify through the informationprovided in the data set.

2.2 Non-Functional Requirements

These kind of behavior shows how the system/module should behave and that it is a constraint upon the systems/module's behavior. Requirements about resource are essential, response time, the accuracy of the data, the effort required to make changes in the software (measurement is personnel effort), reliability.

All of the issues that can be a matter of concern for the storage dedicated to memorize raw data as well as the output of the services, plus the disk storage for the final results is fulfilling the current requirements, since the system can benefit from hard disk.

From a technical aspect all the services are deployed as simple instances that can work in parallel, as well as the rest of the processes. In order to allow many simultaneous users/developers to access the system and set-up/run the services, the platform is delivering the right match of resources in terms of RAM and CPUs.

The platform set up to support the described analytics enables scalable pipelines and processes. The state of the art of the project is implementation phase, relying on a development environment. At this point data is ingested in bulk mode too, in order to collect all the available raw data to archive historical data for the further analysis. The final amount of raw data to ingest and to implement for the use-cases is still to be defined, as well as the system doesn't have specific SLAs to grant.

2.3 Architecture

In this project the first task that we do is to set a dataset containing enough amount of fake news and true news. Then we will do the coding according to the location of the directory at which the

dataset has been stored. Then we use the basic required libraries such as pandas, sklearn where we import them to begin with our coding. Amidst this a TdidfVectorizer will be used to detectthe frequency of the words provided in the data. Later onto the project Machine learningalgorithms will be implemented to the codes so as to verify the accuracy of the provided datasets.And hence the user is now allowed to give in the input with which they will be able to extract thedata with respect to its authenticity as to whether its Fake or Real.

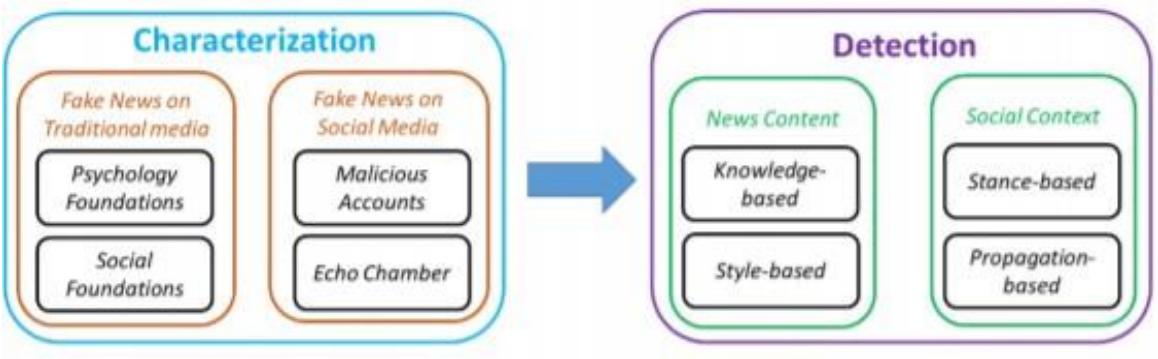

Fig.2.3(a) Characterization and detection of Fake news detection

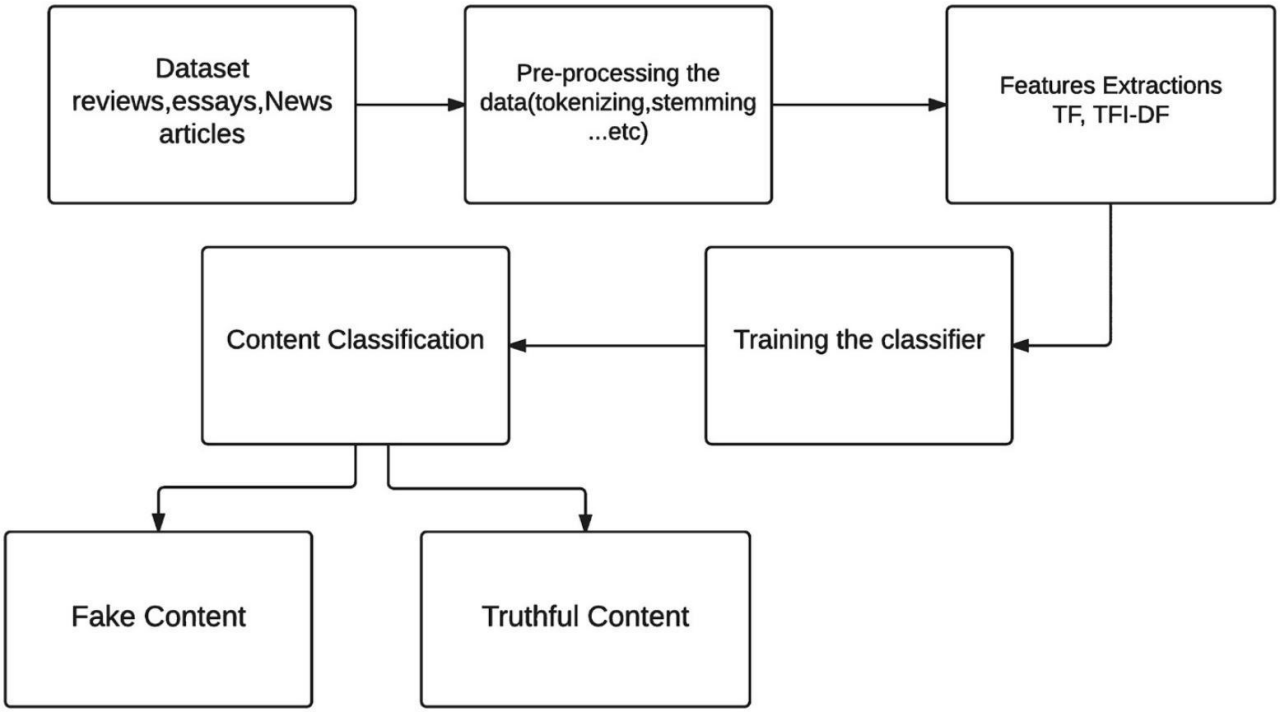

Fig.2.3(b) Architecture of fake news detection

2.4 Flowchart to represent how the processing of fake news works

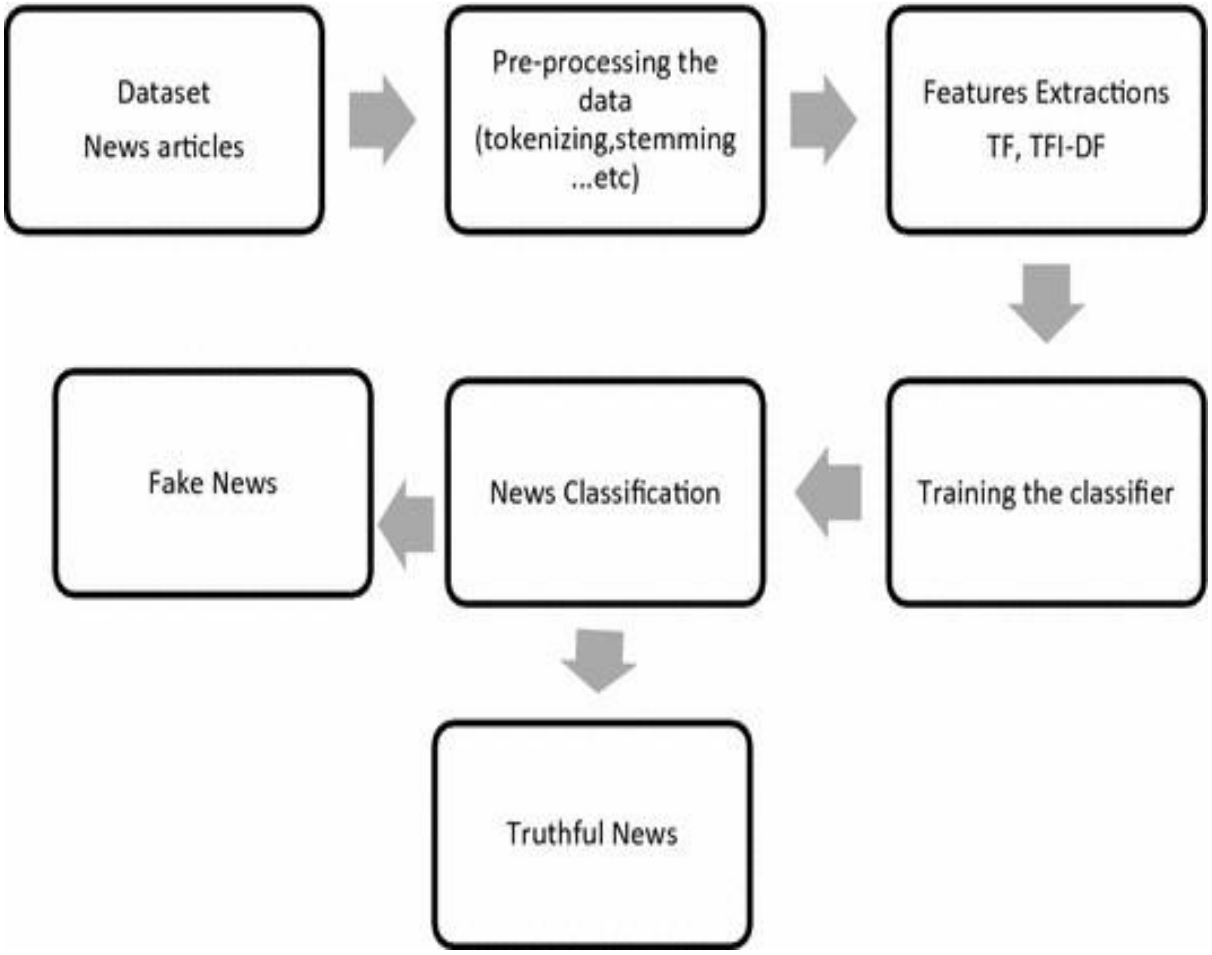

Fig.2.4 Flowchart to represent how the processing of fake news works

2.5. Diagrammatic Representation of classification of news

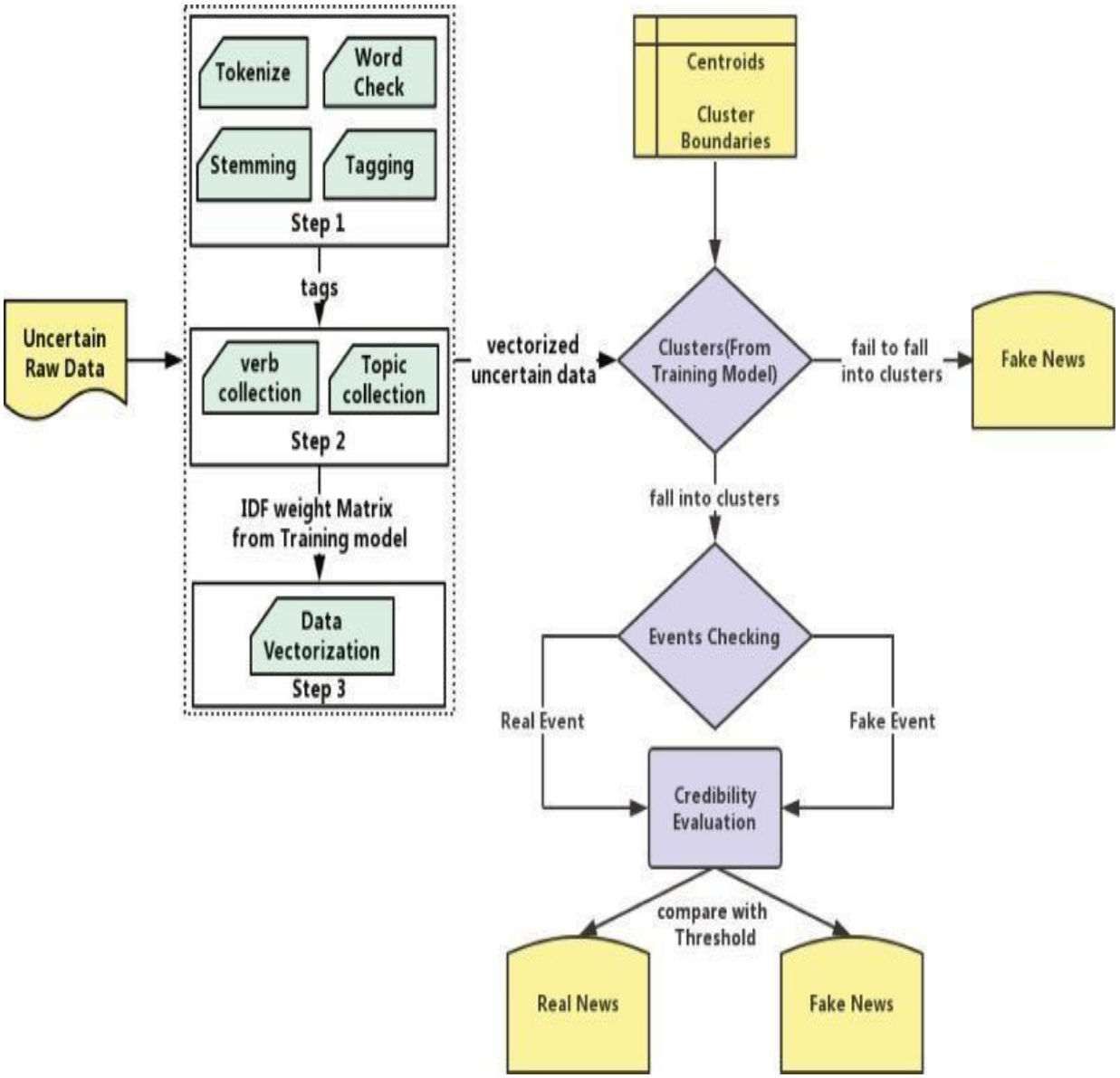

Fig.2.5 Diagrammatic Representation of classification of news into fake news and real news after undergoing certain process

CHAPTER 3

IMPLEMENTATION

3.1 Modules Description

In this project we have 4 Machine Learning classification algorithms and they are :

(1) LINEAR REGRESSION

Machine learning, more specifically the field of predictive modelling is primarily concerned with minimizing the error of a model or making the most accurate predictions possible, at the expense of an explanation. In applied machine learning we will borrow, reuse and steal algorithms from many different fields, including statistics and use them towards these ends. As such, linear regression was developed in the field of statistics and is studied as a model for understanding the relationship between input and output numerical variables, but has been borrowed by machine learning. It is both a statistical algorithm and a machine learning algorithm. Linear regression is a linear model, e.g. a model that assumes a linear relationship between the input variables (x) and the single output variable (y). More specifically, that y can be calculated from a linear combination of the input variables (x). When there is a single input variable (x), the method is referred to as simple linear regression. When there are multiple input variables, literature from statistics often refers to the method as multiple linear regression. Different techniques can be usedto prepare or train the linear regression equation from data, the most common of which is called Ordinary Least Squares. It is common to therefore refer to a model prepared this way as Ordinary Least Squares Linear Regression or just Least Squares Regression. Linear Regressionis an attractive model because the representation is so simple. The representation is a linear equation that combines a specific set of input values (x) the solution to which is the predicted output for that set of input values (y). As such, both the input values (x) and the output value are numeric. The linear equation assigns one scale factor to each input value or column, called a coefficient and represented by the capital Greek letter Beta (B). One additional coefficient is alsoadded, giving the line an additional degree of freedom (e.g. moving up and down on a two-dimensional plot) and is often called the intercept or the bias coefficient.

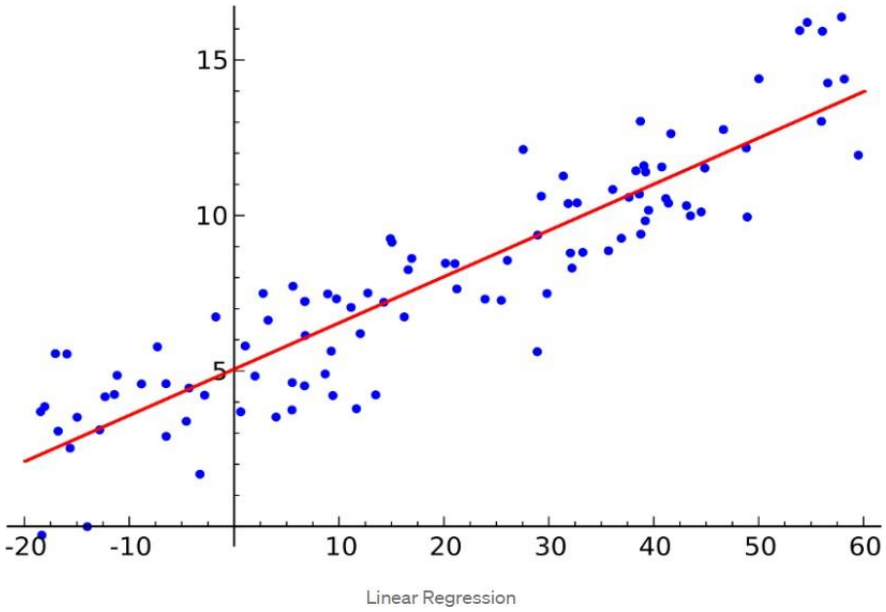

Fig.3.1.1 Linear Regression Graph

(2) DECISION TREE CLASSIFICATION

Decision tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label. A tree can be learned by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions. The construction of decision tree classifier does not require any domain knowledge or parameter setting, and therefore is appropriate for exploratory knowledge discovery. Decision trees can handle high dimensional data. In general decision tree classifier has good accuracy. Decision tree induction is a typical inductive approach to learn knowledge on classification.

Below are some assumptions that we might make while using decision tree:

- At the beginning, we consider the whole training set as the root.
- Feature values are preferred to be categorical. If the values are continuous then they are discretized prior to building the model.
- On the basis of attribute values records are distributed recursively.
- We use statistical methods for ordering attributes as root or the internal node.

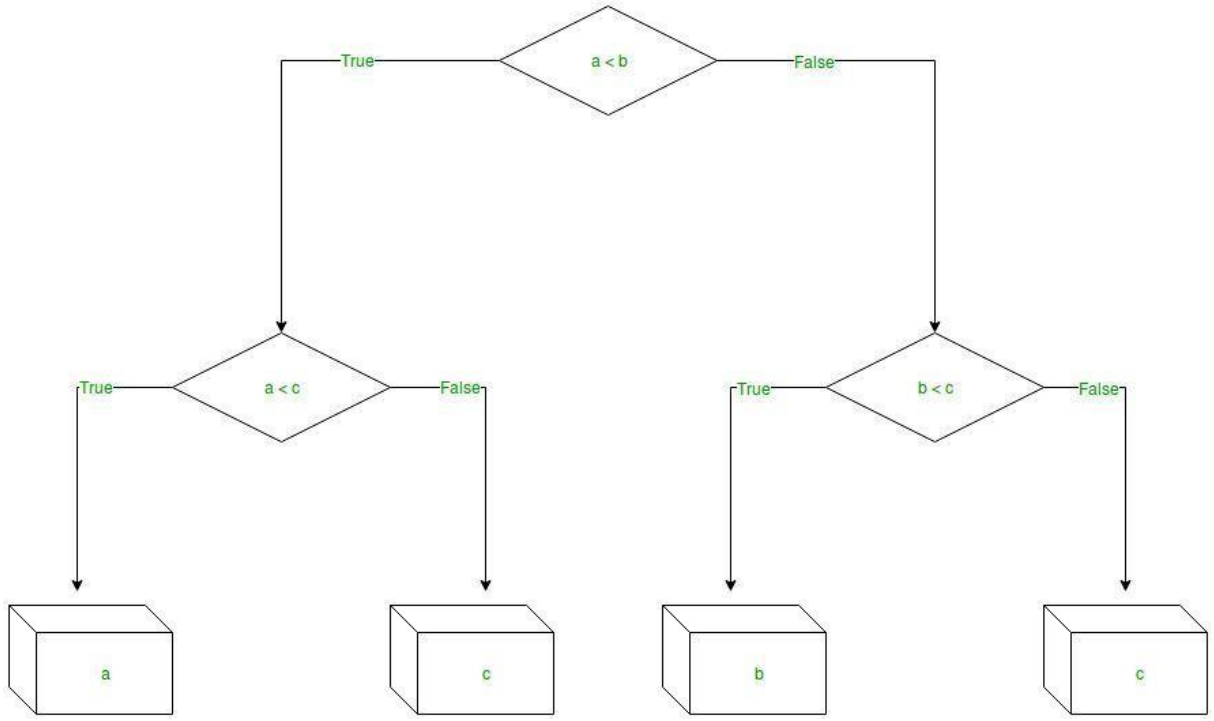

Fig.3.1.2 Decision tree classification

(3) GRADIENT BOOST CLASSIFICATION

Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees.
Gradient boosting re-defines boosting as a numerical optimization problem where the objective is to minimize the loss function of the model by adding weak learners using gradient descent. Gradient Boosting is a first-order iterative optimization algorithm for finding a local minimum of a differentiable function. As gradient boosting is based on minimizing a lossfunction, different types of loss functions can be used resulting in a flexible technique that can beapplied to regression, multi-class classification, etc. Intuitively, gradient boosting is a stage-wise additive model that generates learners during the learning process (i.e., trees are added one at a time, and existing trees in the model are not changed). The contribution of the weak learner tothe ensemble is based on the gradient descent optimization process. The calculated contribution of each tree is based on minimizing the overall error of the strong learner. Gradient boosting doesnot modify the sample distribution as weak learners train on the remaining residual errors of a strong learner (i.e., pseudo-residuals). By training on the residuals of the model, this is an

alternative means to give more importance to misclassified observations. Intuitively, new weak learners are being added to concentrate on the areas where the existing learners are performing poorly. The contribution of each weak learner to the final prediction is based on a gradient optimization process to minimize the overall error of the strong learner.

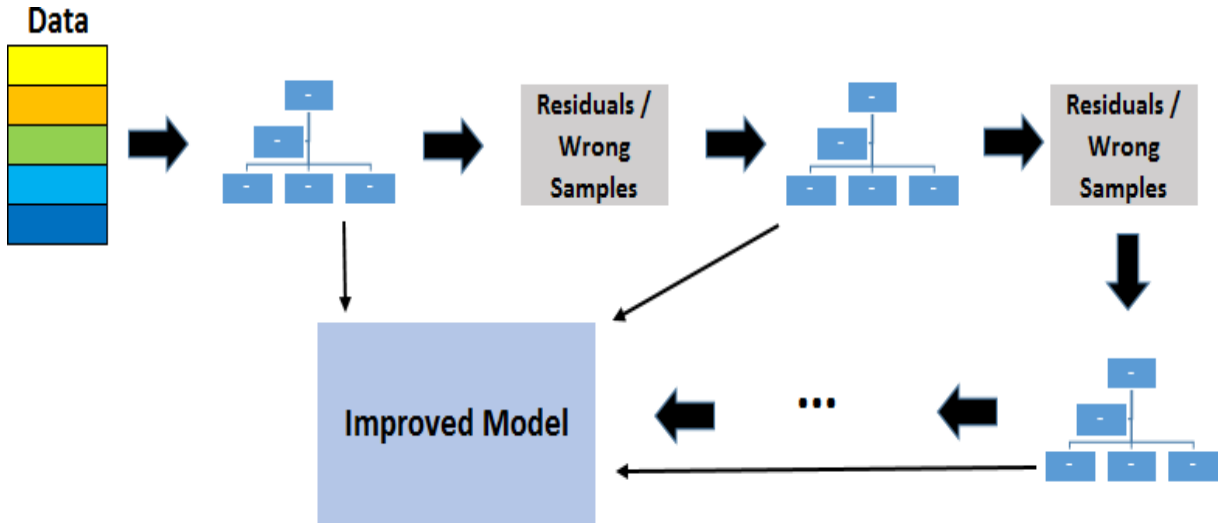

Fig.3.1.3 Gradient boost classification

(4) RANDOM FOREST CLASSIFICATION MODEL

Random forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because of its simplicity and diversity (it can be used for both classification and regression tasks).

Random forest is a supervised learning algorithm. The "forest" it builds, is an ensemble of decision trees, usually trained with the “bagging” method. The general idea of the bagging method is that a combination of learning models increases the overall result. Random forest has nearly the same hyperparameters as a decision tree or a bagging classifier. Fortunately, there'sno need to combine a decision tree with a bagging classifier because you can easily use the classifier-class of random forest. With random forest, you can also deal with regression tasks by using the algorithm's regressor. Random forest adds additional randomness to the model, while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model. Therefore, in random forest, only a random subset of the features is taken into consideration by the algorithm for splitting a node. You can even make trees more random by additionally using random thresholds for each feature rather than searching for the best possible thresholds (like a normal decision tree does).

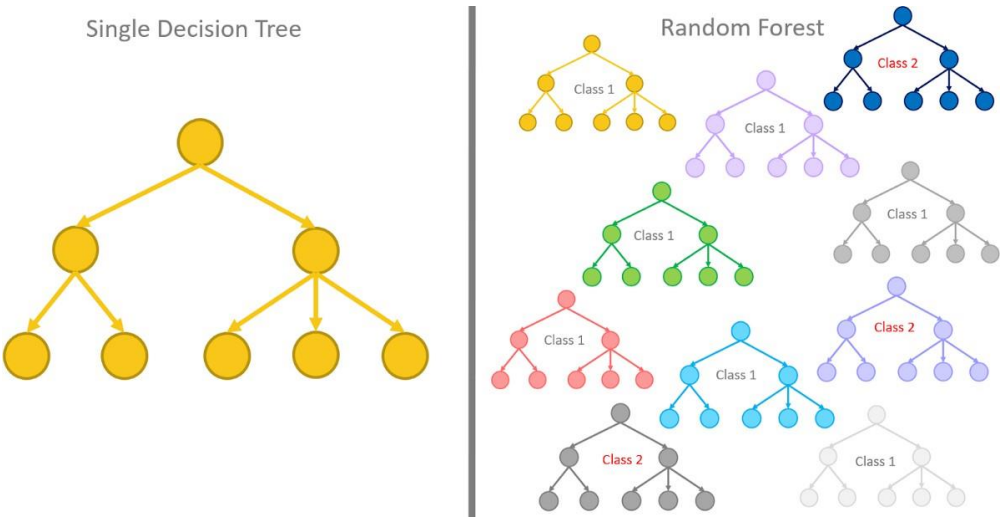

Fig.3.1.4(a) Random Forest classification mode

USED PYTHON PACKAGES :

(1) sklearn :

a. In python, sklearn is a machine learning package which include a lot of ML algorithms.
b. Here, we are using some of its modules like train_test_split, Decision Tree Classifier and accuracy_score.

(2) NumPy :

a. It is a numeric python module which provides fast math functions for calculations.
b. It is used to read data in numpy arrays and for manipulation purpose.

(3) Pandas :

a. Used to read and write different files.
b. Data manipulation can be done easily with datasets.

3.2 Implementation Details

(1) Collect the data set and save it in the required directory.

(2) Launch jupyterlab through anaconda prompt.

(3) Import all the modules required for extracting details for detection of fake news.

(4) Analyze the data and use the TfidfVectorizer to detect the frequency of words.

(5) Machine learning algorithms such as Linear Regression, Decision Tree classification, Gradient boost classification and Random forest classification are used to detect the accuracy of the dataset.

(6) Lastly the user can check the authenticity of the provided information in the data set as to whether it is real or fake.

(7) Hence, the project is created.

3.3 TOOLS USED

(1) Python

(2) Conda

(3) jupyterLab

(4) sklearn

(5) NumPy

(6) Pandas

CHAPTER 4

TEST RESULTS/VERIFICATION/RESULTS

4.1 TESTING

Two datasets containing fake news and true news will be saved in the directory during the initial part of doing the project.

(A) True news

	A	B	C	D	E
1	title	text	subject	date	
2	As U.S. bu	WASHING	politicsNe	December 31, 2017	
3	U.S. milita	WASHING	politicsNe	December 29, 2017	
4	Senior U.S	WASHING	politicsNe	December 31, 2017	
5	FBI Russia	WASHING	politicsNe	December 30, 2017	
6	Trump wa	SEATTLE/V	politicsNe	December 29, 2017	
7	White Hou	WEST PALI	politicsNe	December 29, 2017	
8	Trump say	WEST PALI	politicsNe	December 29, 2017	
9	Factbox: T	The follow	politicsNe	December 29, 2017	
10	Trump on	The follow	politicsNe	December 29, 2017	
11	Alabama c	WASHING	politicsNe	December 28, 2017	
12	Jones certi	(Reuters) -	politicsNe	December 28, 2017	
13	New York	NEW YORI	politicsNe	December 28, 2017	
14	Factbox: T	The follow	politicsNe	December 28, 2017	
15	Trump on	The follow	politicsNe	December 28, 2017	
16	Man says I	(In Dec. 2	politicsNe	December 25, 2017	
17	Virginia of	(Reuters) -	politicsNe	December 27, 2017	
18	U.S. lawm	WASHING	politicsNe	December 27, 2017	
19	Trump on	The follow	politicsNe	December 26, 2017	
20	U.S. appe	(Reuters) -	politicsNe	December 26, 2017	
21	Treasury S	(Reuters) -	politicsNe	December 24, 2017	
22	Federal juc	WASHING	politicsNe	December 24, 2017	
23	Exclusive:	NEW YORI	politicsNe	December 23, 2017	
24	Trump tra	(Reuters) -	politicsNe	December 23, 2017	
25	Second co	WASHING	politicsNe	December 23, 2017	
26	Failed vote	LIMA (Reu	politicsNe	December 23, 2017	
27	Trump sigr	WASHING	politicsNe	December 22, 2017	

Fig.4.1(A) True news dataset

(B) Fake news

	title	text	subject	date
2	Donald Tr	Donald Tru	News	December 31, 2017
3	Drunk Bra	House Inte	News	December 31, 2017
4	Sheriff Da	On Friday,	News	December 30, 2017
5	Trump Is S	On Christn	News	December 29, 2017
6	Pope Fran	Pope Franc	News	December 25, 2017
7	Racist Ala	The numbe	News	December 25, 2017
8	Fresh Off	Donald Tru	News	December 23, 2017
9	Trump Sai	In the wak	News	December 23, 2017
10	Former CI	Many peop	News	December 22, 2017
11	WATCH: E	Just when	News	December 21, 2017
12	Papa John	A centerpi	News	December 21, 2017
13	WATCH: P	Republicar	News	December 21, 2017
14	Bad News	Republicar	News	December 21, 2017
15	WATCH: L	The media	News	December 20, 2017
16	Heiress To	Abigail Dis	News	December 20, 2017
17	Tone Dea	Donald Tru	News	December 20, 2017
18	The Interr	A new anir	News	December 19, 2017
19	Mueller Sp	Trump sup	News	December 17, 2017
20	SNL Hilari	Right now,	News	December 17, 2017
21	Republica	Senate Ma	News	December 16, 2017
22	In A Heart	It almost s	News	December 16, 2017
23	KY GOP St	In this #ME	News	December 13, 2017
24	Meghan M	As a Demo	News	December 12, 2017
25	CNN CALL	Alabama is	News	December 12, 2017
26	White Hou	A backlash	News	December 12, 2017
27	Despicabl	Donald Tru	News	December 12, 2017

Fig.4.1(B) Fake news dataset

1. Logistic Regression

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

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

```
LogisticRegression()
```

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

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

```
0.9885026737967915
```

Fig.4.1.1 Logistic Regression

2. Decision Tree Classification

```
from sklearn.tree import DecisionTreeClassifier
```

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

```
DecisionTreeClassifier()
```

```
pred_dt = DT.predict(xv_test)
```

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

```
0.9945632798573975
```

Fig.4.1.2 Decision tree classification

3. Gradient Boosting Classifier

```
from sklearn.ensemble import GradientBoostingClassifier
```

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

```
GradientBoostingClassifier(random_state=0)
```

```
pred_gbc = GBC.predict(xv_test)
```

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

```
0.9955436720142602
```

Fig.4.1.3 Gradient boosting classifier

4. Random Forest Classifier

```
from sklearn.ensemble import RandomForestClassifier
```

```
RFC = RandomForestClassifier(random_state=0)
RFC.fit(xv_train, y_train)
```

```
RandomForestClassifier(random_state=0)
```

```
pred_rfc = RFC.predict(xv_test)
```

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

```
0.9915329768270945
```

Fig.4.1.4 Random Forest classifier

The accuracy level of the datasets has been successfully verified by the above concepts that are mentioned and the accuracy score is pretty much good.

4.2 Verification

After the above-mentioned testing is done, we proceed further for the manual testing wherein theuser can input the details and know whether that particular news is fake or not.

Model Testing With Manual Entry

News

```
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_DT = DT.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: {}".format(output_lable(pred_LR[0]),
                                                                                                              output_lable(pred_DT[0]),
                                                                                                              output_lable(pred_GBC[0]),
                                                                                                              output_lable(pred_RFC[0])))
```

```
news = str(input())
manual_testing(news)
```

Fig.4.2Verification (Input)

```
Williams (shadowfacts) reports, the unemployment rate that includes those Americans who have given up looking for a job because there are
no jobs to be found is 23%.The Federal Reserve, a tool of a small handful of banks, has succeeded in creating the illusion of an economic
recovery since June, 2009, by printing trillions of dollars that found their way not into the economy but into the prices of financial as
sets.  Artificially booming stock and bond markets are the presstitute financial media s  proof  of a booming economy.The handful of lear
ned people that America has left, and it is only a small handful, understand that there has been no recovery from the previous recession
and that a new downturn is upon us.  John Williams has pointed out that US industrial production, when properly adjusted for inflation, h
as never recovered its 2008 level, much less its 2000 peak, and has again turned down.The American consumer is exhausted, overwhelmed by
debt and lack of income growth. The entire economic policy of America is focused on saving a handful of NY banks, not on saving the Ameri
can economy.Economists and other Wall Street shills will dismiss the decline in industrial production as America is now a service econom
y. Economists pretend that these are high-tech services of the New Economy, but in fact waitresses, bartenders, part time retail clerks,
and ambulatory health care services have replaced manufacturing and engineering jobs at a fraction of the pay, thus collapsing effective
aggregate demand in the US. On occasions when neoliberal economists recognize problems, they blame them on China.It is unclear that the U
S economy can be revived. To revive the US economy would require the re-regulation of the financial system and the recall of the jobs and
US GDP that offshoring gave to foreign countries. It would require, as Michael Hudson demonstrates in his new book, Killing the Host, a r
evolution in tax policy that would prevent the financial sector from extracting economic surplus and capitalizing it in debt obligations
paying interest to the financial sector.The US government, controlled as it is by corrupt economic interests, would never permit policies
that impinged on executive bonuses and Wall Street profits.  Today US capitalism makes its money by selling out the American economy and
the people dependent upon it.In  freedom and democracy  America, the government and the economy serve interests totally removed from the
interests of the American people. The sellout of the American people is protected by a huge canopy of propaganda provided by free market
economists and financial presstitutes paid to lie for their living.When America fails, so will Washington s vassal states in Europe, Cana
da, Australia, and Japan.  Unless Washington destroys the world in nuclear war, the world will be remade, and the corrupt and dissolute W
est will be an insignificant part of the new world.Dr. Paul Craig Roberts was Assistant Secretary of the Treasury for Economic Policy and
associate editor of the Wall Street Journal. He was columnist for Business Week, Scripps Howard News Service, and Creators Syndicate. He
has had many university appointments. His internet columns have attracted a worldwide following. Roberts  latest books are The Failure of
Laissez Faire Capitalism and Economic Dissolution of the West, How America Was Lost, and The Neoconservative Threat to World Order. READ
MORE NWO NEWS AT: 21st Century Wire NWO Files


LR Prediction: Fake News
DT Prediction: Fake News
GBC Prediction: Fake News
RFC Prediction: Fake News
```

Fig.4.2 Verification (Output)

So here in the above provided image you can see the output for which the input was provided by the user that is me. So now the machine learning concepts have verified the data thoroughly and has undergone so many algorithmic verifications wherein the result provided was that the data is absolutely fake which means that it is a fake
news.

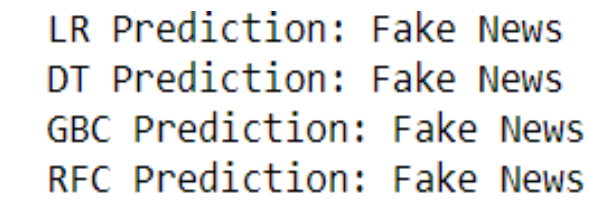

Fig.4.2 Result showing fake news

4.3 Result

The project on Detection of Fake news has been executed successfully using python programming language.
Machine learning algorithms have been used to verify the data accuracy and hence the results were the proven fact for it to have an excellent accuracy rate.
After the implementation of the above-mentioned processes a manual testing was done wherein the user was supposed to input the data and hence the result proved to provide the accurate information about its authenticity as to whether it is a fake news or real news. Then upon checking the news seemed to be absolutely fake.
Hence the project Fake News Detection has been implemented, executed and verified successfully.

CHAPTER 5
CONCLUSIONS AND FURTHER SCOPE

5.1 Conclusions:

- For this project, I took the online available Data sets and using python programming analyzed and processed the data. After that the data had to undergo certain verification processes to check the frequency of its words and the accuracy of the data which was checked by machine learning algorithms. And then I used this knowledge to build the best model for detecting fake news using manual testing.

- This project further helped me in understanding Python, its libraries and machine learning concepts. This knowledge that I have acquired through this project will for sure help me in future to further develop more and more projects.

5.2 FURTHER SCOPE:

- Create a separate model for each genre of news that is spread worldwide.
- Create new features from various other complicated version of tools.
- Furthermore, data sets can also be added which will become a massive dataset and fake newscan be detected through it.
- This kind of project can also be used in certain organizations to detect the spread of false information that will help in not misleading the clients.

REFERENCES

Wikipedia:

- https://en.wikipedia.org/wiki/Machine_learning
- https://www.geeksforgeeks.org/best-python-libraries-for-machine-learning/
- https://en.wikipedia.org/wiki/Detecting_fake_news_online

Websites:

- https://jupyter.org/
- https://anaconda.org/

Books:

- Python Crash course: A Hands-on Project Based Programming by Eric Matthews.
- The Hundred-Page Machine Learning Book by Andriy Burkov
- Machine Learning yearning by Andrew N