

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY

NOIDA

**FUNDAMENTALS OF MACHINE LEARNING**

**PROJECT BASED LEARNING**

**PROJECT NAME:**

**FAKE NEWS DETECTION USING PYTHON MACHINE LEARNING**

**SUBMITTED TO:**

**MS. PARUL AGARWAL**

**SUBMITTED BY:**

**ANURAG SINGH 19103150**

**SHIVANK GOYAL 19103124**

**BATCH B4**

**INDEX**

1. TITLE
2. PROBLEM STATEMENT
3. MOTIVATION
4. LITERATURE AVAILABLE ON THE PROBLEM
5. MODEL USED
6. FLOW CHART
7. IMPLEMENTATION
8. RESULTS
9. CONCLUSION
10. REFERENCES

**Problem Statement:**

In this project we have created a **machine learning model** to detect the news whether the news is fake or real. The Internet is an incredible resource for news and information, but unfortunately not everything online is trustworthy. **Fake news** is any article or video containing untrue information disguised as a credible news source. While fake news is not unique to the Internet, it has recently become a big problem in today’s digital world.

To detect the news we have used TfidVectorizer to detect the news, TF will check the frequency of the terms and IDF is a measure of how significant a term is in the entire corpus.

**Motivation:**

This project is very useful to detect the real world statements as real or fake. Most of the news which circulates on the internet are fake only some of them are real. This model will easily classify the news as fake or real. This model is made using Python one the best growing language in the field of machine learning or model/machine training. We have arranged the real data of fake and real news of approximately 23000 data entries in each file.

**Literature available on the problem:**

Internet is one of the important inventions and a large number of persons are its users. These persons use this for different purposes. There are different social media platforms that are accessible to these users. Any user can make a post or spread the news through these online platforms. These platforms do not verify the users or their posts. So some of the users try to spread fake news through these platforms. These fake news can be a propaganda against an individual, society, organization or political party. A human being is unable to detect all these fake news. So there is a need for machine learning classifiers that can detect these fake news automatically. Use of machine learning classifiers for detecting the fake news is described in this systematic literature review.

**Model Used:**

Machine Learning Model: A machine learning model is a file that has been trained to recognize certain types of patterns. You train a model over a set of data, providing it an algorithm that it can use to reason over and learn from those data.

Libraries used:

1. Pandas: Pandas is built on top of two core Python libraries—matplotlib for data visualization and NumPy for mathematical operations.
2. NumPy: N**umPy** offers comprehensive mathematical functions, random number generators, linear algebra routines, Fourier transforms, and more.
3. Sklearn: Scikit-learn is probably the most useful library for machine learning in Python. The sklearn library contains a lot of efficient tools for machine learning and statistical modelling including classification, regression, and clustering and dimensionality reduction.
4. TfidVectorizer: TF-IDF is an abbreviation for Term Frequency Inverse Document Frequency. This is very common algorithm to transform text into a meaningful representation of numbers which is used to fit machine algorithm for prediction.

Language: PYTHON

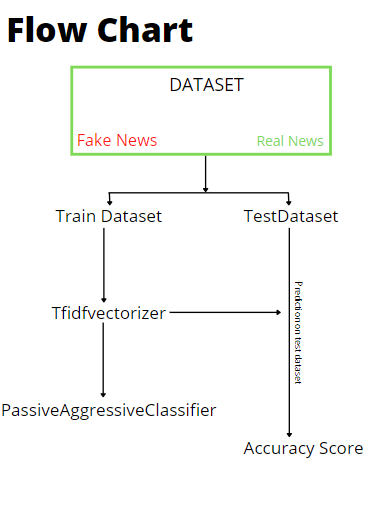
Software:

1. JUPYTER NOTEBOOK
2. Anaconda prompt

Accuracy calculated using:

1. Logistic Regression: Accuracy is the proportion of correct predictions over total predictions.
2. Decision Tree: The number of correct predictions made divided by the total number of predictions made.
3. Random Forest Classifier: 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.

**Flow Chart:**

****

**Implementation:**

1. Importing all the required libraries and the dataset of both fake and real news, after visualising the data removing all the data which is not required to predict the news whether the news is real or fake.
2. Dividing last few rows of both the dataset into a new dataset to test the model at the end for manual testing.
3. Splitting the data into test and train dataset using train\_test\_split function of Sklearn.
4. Using TfidfVectorizer to convert a collection of raw documents to a matrix of TF-IDF features. TF stands for term frequency which means number of times a word appears in a document is its Term Frequency. A higher value means a term appears more often than others, and so, the document is a good match when the term is part of the search terms.

IDF stands for Inverse Document Frequency:

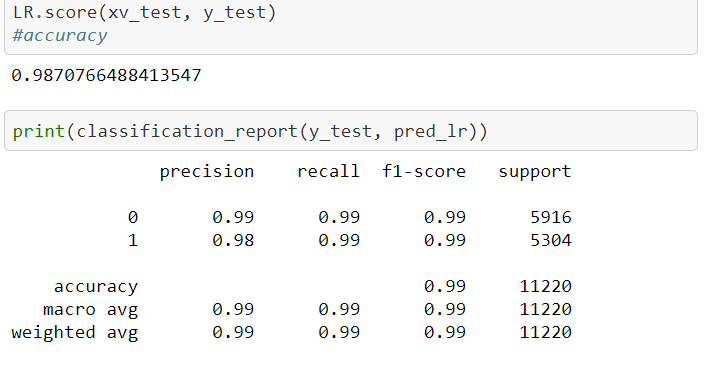
The TfidfVectorizer converts a collection of raw documents into a matrix of TF-IDF features.

1. Using Passive Aggressive Classifier Algorithms to correct classification outcome, and turns aggressive in the event of a miscalculation, updating and adjusting.
2. Applying Logistic Regression, Decision tree classifier and Random forest on test and train dataset to get the accuracy score for each method.
3. At the end adding a function of check the input news statement is fake or not using this model.

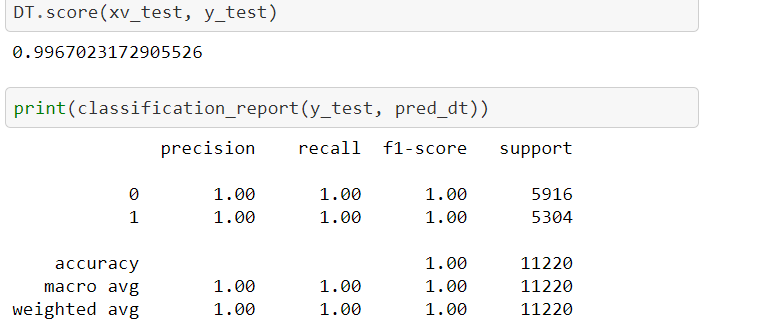
Code link: <https://github.com/anuragc10/FakeNewsDetection/blob/main/Fake%20news%20Detection%20github.ipynb>

**Results:**

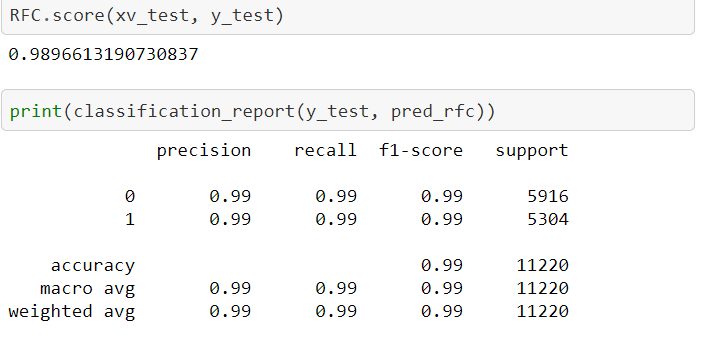
1. Logistic Regression Accuracy Score**: 0.9870766488413547**



1. Decision Tree Accuracy Score**: 0.9967023172905526**



1. Random Forest Accuracy Score**: 0.9896613190730837**



**Input News:** Just when you might have thought we d get a break from watching people kiss Donald Trump s ass and stroke his ego ad nauseam, a pro-Trump group creates an ad that s nothing but people doing even more of those exact things. America First Policies is set to release this ad, called Thank You, President Trump, on Christmas Day and, well, we threw up a little in our mouths trying to watch this.Basically, the spot is nothing but people fawning all over Trump for all the stuff he hasn t actually done. The ad includes a scene with a little girl thanking Trump for bringing back Merry Christmas, which never went away (there are even videos of President Obama saying Merry Christmas himself). A man thanks him for cutting his taxes. And America First says that everyday Americans everywhere are thanking Trump for being such a great and awesome president.The best president.Nobody s ever done what he s done. He s breaking all kinds of records every day.Believe us.Anyway, the word propaganda comes to mind when watching this. That s what it is literal propaganda promoting someone who shouldn t need this kind of promotion anymore. Watch this ad bullshit below:The way the MAGAs are kowtowing to Orange Hitler is both disgusting and frightening. The man has done nothing, and his policies will harm the very same Americans who are thanking him. Unfortunately, it will take an obscene amount of pain before they ll open their eyes and see they ve been duped by a con man with a bad hairdo.And his ongoing need for this kind of adoration is, at best, unbecoming of his office. This ad is vile.Featured image via Al Drago-Pool/Getty Images

**Output:**

Logistic regression Prediction: Fake News

Decision tree Prediction: Fake News

Random forest Prediction: Fake News

**Conclusion:**

The purpose of this machine learning model is to identify out the news statement is fake or real. Our machine learning model is upto 99 percent accurate by considering 3 accuracy scores.

This model can easily figure out the news if it is real or fake. This model can be used in many different ways and could improve the situations by checking the news. This project helps us a lot to learn new technologies and their real time uses. We have tried our best to make this model more easy to understand and accurate.

**References:**

Kaggle.com

Wikipedia.com

Python.org

Jupyter.org