**Approach**

Fake News Detection Using Python And

Machine Learning

Team Name: Tricoders

Team Members:

1. Maanickavasagan P
2. Kethuram T A
3. Kiranvignesh S

Steps undertaken to achieve the solution for the given problem statement:

1)Data Collection

2)DataPreprocessing

3)Train and Test Split

4)Text Vectorization

5)Model Selection and Training

6)Testing the model and Prediction with example

7)Evaluation of model

1.**Data Collection:**

Data Collection is the process that involves preparing/collecting the dataset which is going to be utilized in the

model. The dataset to be collected varies for each model, based on the needs and constraints.

Reasoning:

* The information required to deliver answers, analyze business performance or other outcomes, and forecast future trends, actions, and scenarios is provided by effective data collecting.
* The ISOT Fake News Dataset is the one gathered here.

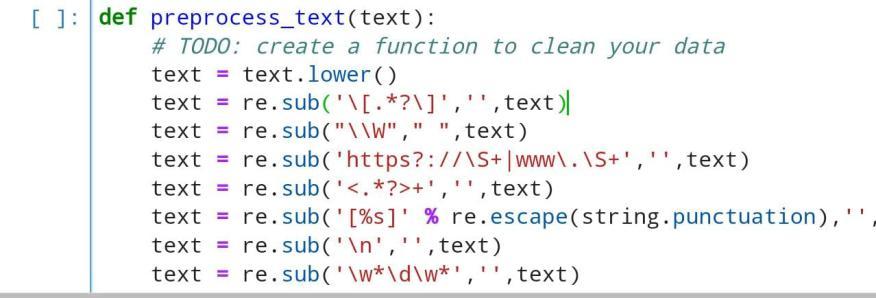
Outcome:The ISOT Fake news Dataset is downloaded and uploaded into the data directory in the Intel JupyterLab platform.

1. **Data Preprocessing:**

Data preprocessing is the process of putting raw datasets into a format that may be used, consumed, and understood for more in-depth study.

Examples of data preprocessing include cleaning, instance selection, normalization, transformation, feature extraction and selection.

Data Preprocessing:



Reasoning:

Before using the algorithm, the dataset is preprocessed to look for missing values, noisy data, and other irregularities.

.

Outcome:

The data which are useless in the dataset has been cleared from the dataset and the dataset has been cleaned.Changes are applied to the dataset.

3.**Train and Test Split:**

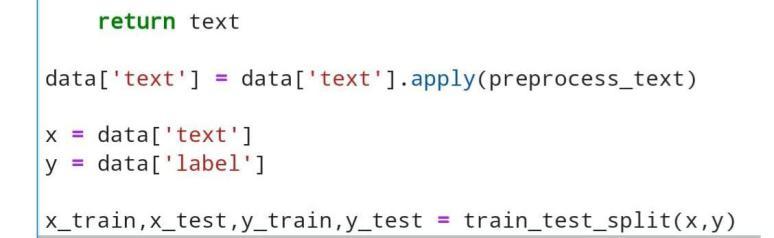
The train/test method is a means to evaluate your model's accuracy. The data set was given the name "Train/Test" because you split it into two sets, a training set and a testing set.

Reasoning:

* To prevent any issues with overfitting and underfitting, data is divided into train and test sets.
* To evaluate prediction performance objectively, you must divide your dataset.

Outcome:

The whole dataset is divided into two sections where 80% of original dataset is allocated for training and 20% is allocated for testing.



4.**Text Vectorization:**

The process of transforming text representation into numerical representation is called text vectorization.

The following list of popular techniques includes:

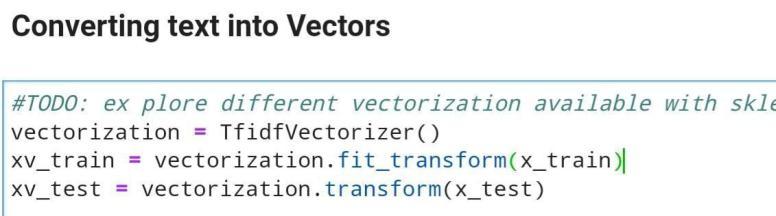
* Binary Term Frequency
* Bag of Words (BoW) Term Frequency
* (L1) Normalized Term Frequency
* (L2) Normalized TF-IDF
* Word2Vec

Reasoning:

Vectorization enables us to avoid such “for” loops, and use all the data in one step.The idea is to get some distinct features out of the text for the model to train on, by converting text to numerical vectors.TF-IDF Vectorizer,where it converts the words to numbers based on Total Frequency and Inverse Document Frequency of words.

Outcome:

The TF-IDF Vectorizer is used to the news to convert it into numbers.



5.**Model Selection and Training:**

In this we select the suitable machine learning model to get the required solution.The models that we can use are

* linear SVC
* Naive Bayes
* Logistic regression
* Random forest

These models are well suited.

Reasoning:

The model is trained using the Logistic Regression algorithm. It is a supervised machine learning algorithm which is utilized for binary classification.

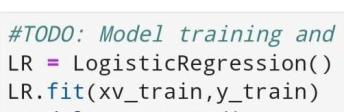
We classify the news into two categories fake and true news.So we can use Logistic Regression algorithm.

By training the model, the model can understand the relationship between the independent and dependent variables and have the ability to make future predictions

Outcome:

The model is trained by Logistic Regression Successfully.

We train the model so the model can understand the relationship between the independent and dependent variables and have the ability to make future predictions.



6.**Testing the Model and Prediction with example:**

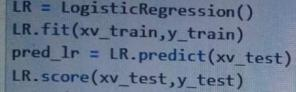
Testing is the process when a fully trained model's performance is assessed using a testing set.

Reasoning:

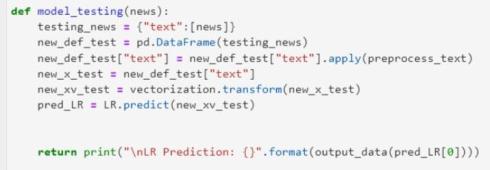
After the model has successfully been trained, we test it using a testing dataset to compare the model's predictions with the actual outcomes.

Outcome:

The results of the model's testing serve as the basis for its appraisal.



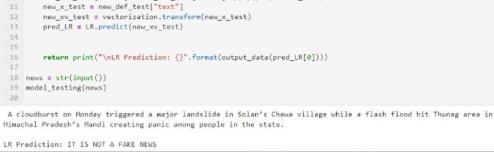
Then we test and predict the model with some example news,



Testing with the Fake News:



Testing with not a Fake News:



7.**Evaluation of Model:**

Model evaluation is a procedure that uses measures to assess the performance of the model, including accuracy, F1 score, recall, and precision.

Reasoning:

When conducting preliminary research, it's critical to evaluate models to determine their effectiveness. Finding the model that best depicts our data and predicts how well the model will perform in the future is helpful.

Outcome:

