**DOCUMENTATION OF FAKE NEWS DECTECTOR**

**IDE** -JUPYTER NOTEBOOK

STEP 1: DATA COLLECTION

**1. Source**- Kaggle " https://www.kaggle.com/datasets/emineyetm/fake-news-detection-datasets ".

**2. Number of data points** -44941

TRUE data points- 21438

FLASE datapoints-23503

-COMBINED BOTH DATASETS INTO ONE

3. **FEATURES**- title, text, subject, data, type(1/0)

>TITLE- TITLE OF NEWS

>TEXT- DESCRIPTION OF THE NEWS

>SUBJECT- Shows Type of News

>DATA - Shows date of news

>TYPE- shows news is true or flase news(1(true) and 0(flase))

**4. Final data**- This is the final dataset which I want to use further.

STEP 2 :**Data Exploration and Data Preprocessing**

* Data Exploration and Cleaning:

1. Loaded dataset into data frame using pandas libraries .

2 .data exploration done and understood its datatypes and description of data.

3. Removed unnecessary columns from the data set .

4. Removal of null values and duplicates - removal of 6k duplicate values done and no null values found.

5. Known how to deal with missing values (can use mean, mode, median and knnimputer to impute data.

6. Plotted histogram , bar graph and pie chart after cleaning the data- comparison between true and false news dataset is done via these graph.

7. Checked weather the data is balances or not via pie chart.

CURRENT DATASET RECORDS- 38657

PRESENT COLUMNS IN USE- TITLE,TEXT,TYPE(LABEL 0/1)

* **Pre-processing: Used NLTK library.**

1. Tokenization is done.
2. Lowercasing: Convert all text to lowercase to ensure uniformity.
3. Removal of punctuation and stop words .
4. Applied stemming and lemmatization for columns .
5. Removed special characters and Symbols.