Dataset Description

Fake News Detection using Python and Machine Learning

Team Name:PSA

Team Members:
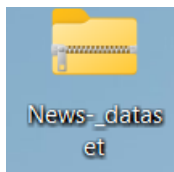
1) M Mohamed Aslam
2) Sahana K
3) Prethisha V

Steps for downloading the dataset:

- Link to the dataset:

  https://onlineacademiccommunity.uvic.ca/isot/2022/11/27/fake-news-detection-datasets/

- We can then download the ISOT Fake News dataset which is stored in a zip folder named "News Dataset"

- This folder consists of our datasets: Fake.csv and True.csv

Fake	07/12/23 8:49 PM	XLS Worksheet	61,319 KB
True	07/12/23 8:49 PM	XLS Worksheet	52,328 KB

- Fake.csv

	title	text	subject	date
1	title	text	subject	date
2	Donald Tru	Donald Trur	News	31-Dec-17
3	Drunk Brag	House Intell	News	31-Dec-17
4	Sheriff Dav	On Friday, it	News	30-Dec-17
5	Trump Is Sc	On Christma	News	29-Dec-17
6	Pope Franc	Pope Franci	News	25-Dec-17

- True.csv

	title	text	subject	date
1	title	text	subject	date
2	As U.S. budg	WASHINGT(	politicsNew	31-Dec-17
3	U.S. military	WASHINGT(	politicsNew	29-Dec-17
4	Senior U.S.	WASHINGT(	politicsNew	31-Dec-17
5	FBI Russia p	WASHINGT(	politicsNew	30-Dec-17
6	Trump want	SEATTLE/W	politicsNew	29-Dec-17

Description:

- The Fake News Dataset consists of 23503 news records,while The True News Dataset consists of 21418 news records,which in total adds up 44921 records.

- Collection of fake and true news datasets are closely balanced(23503 fake news records v/s 21418 true news records)which allows to provide a more balanced result.

- News are obtained from different legitimate sites,which improves model's reliability

- The title and text columns are the most essential features to focus on as these are useful for detecting a news is fake or not.However the subject column which specifies the type of news and date are not significant while detecting the fake news.
- Labels are assigned to indicate the fake and true news in the separate datasets

Assigning Classes to the Dataset

```
[6]: true_data["label"] = 1
     fake_data["label"] = 0
```