OBJECTIVES

1

Develop a tool to identify and classify news articles as fake or genuine

2

Enhance media literacy by promoting critical thinking and skepticism towards online content.
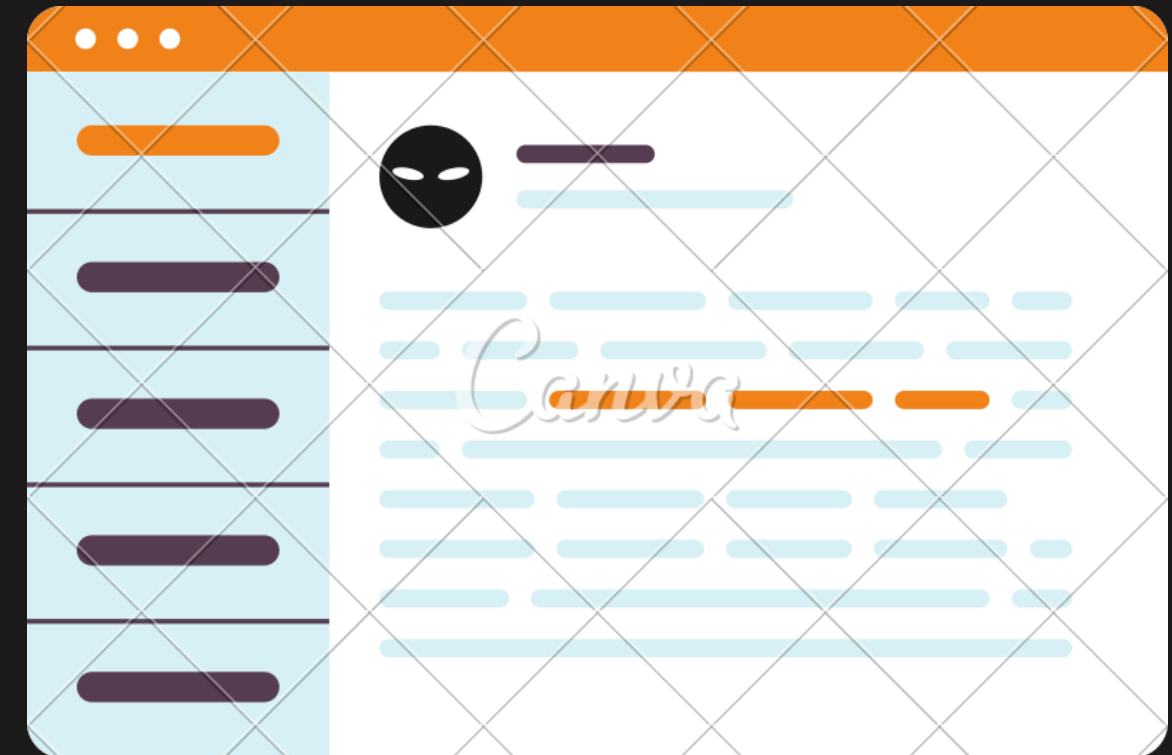
3

Contribute to combating misinformation and promoting credible journalism.

veriFi

Unmasking Truth, Shielding Minds

IMPORTANCE OF FAKE NEWS DETECTOR

- Rising proliferation of misinformation on digital platforms.
- Potential to influence public opinion and societal perceptions.
- Essential for preserving trust in media and democratic discourse.

Think of an email or message you received that asked for personal information. What made it suspicious?

ABOUT DATASET

	title	text	subject	date	label
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0

	text	label
0	WASHINGTON (Reuters) - President Donald Trump ...	1
1	WASHINGTON (Reuters) - U.S. lawmakers released...	1
2	PARIS (Reuters) - French President Emmanuel Ma...	1
3	Donald Trump and his administration have prove...	0
4	The Democratic infighting between Bernie Sande...	0

```
news.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 44898 entries, 0 to 44897
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   text    44898 non-null  object
 1   label   44898 non-null  int64
dtypes: int64(1), object(1)
memory usage: 701.7+ KB
```

LIBRARIES USED

Phishing attacks come in different forms

JACKSON OBJECTMAPPER

- For JSON data binding and deserialization.
- Usage: Used to map JSON file data to Java objects representing news articles.

WEKA

- **J48Decision Tree Classifier:** Used for training a decision tree model on vectorized text data.
- **FilteredClassifier**: Employed to encapsulate a classifier and a filter to process input data.

01 DATA COLLECTION AND PREPROCESSING:

○ **Data Source**: JSON file containing news articles.
○ **ObjectMapper**: Reads JSON file and maps news data to Java objects.
○ **TextPreprocessor**: Cleans and preprocesses textual data for model training and prediction.

02 MODEL BUILDING AND TRAINING

- **ModelBuilder Method**:
  - Loads and preprocesses news data.
  - Creates attribute structures and Instances for training.
  - Applies StringToWordVector filter for feature extraction.
  - Trains a J48 decision tree classifier on vectorized Instances.
  - Evaluates classifier performance using cross-validation.

03 USER INPUT PREDICTION

- **predictFromUserInput Method**:
  - Preprocesses user input text.
  - Creates Instances for user input data.
  - Applies StringToWordVector filter to vectorize the user input data.
  - Classifies the user input instance using the trained model.
  - Returns predicted label (fake or genuine) based on the classification result.

MODEL EVALUATION

```
Number of instances after filtering: 6000
Decision Tree Classifier Evaluation:


Correctly Classified Instances        5941                99.0167 %
Incorrectly Classified Instances        59                 0.9833 %
Kappa statistic                          0.9803
Mean absolute error                      0.0139
Root mean squared error                  0.0969
Relative absolute error                  2.7785 %
Root relative squared error             19.3885 %
Total Number of Instances             6000
```

SAMPLE OUTPUT

veriFi

Unmasking Truth, Shielding Minds

e itself in this election. You want to give me a good sendoff? Go vote! H/T [Breitbart]

Verify

The provided news is Genuine News