

FAKE NEWS DETECTION USING NLP

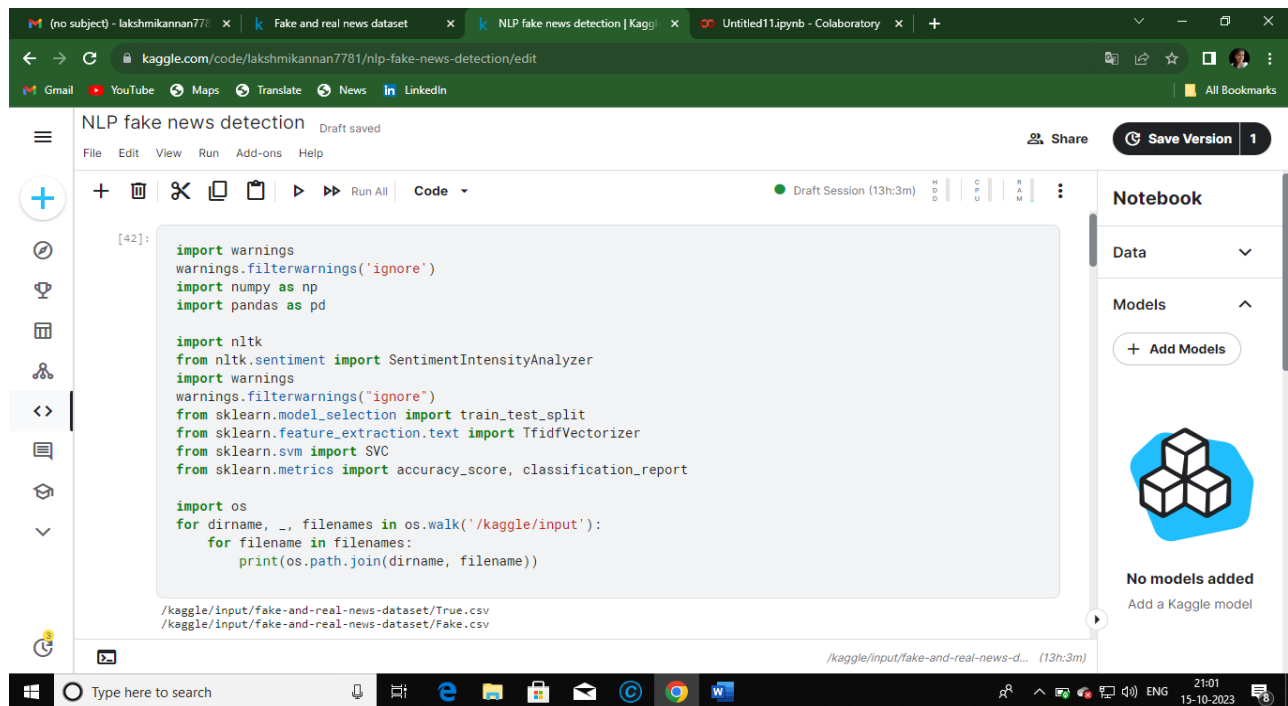
PHASE 3

Loading and Preprocessing the dataset

Dataset link:

<https://www.kaggle.com/datasets/clmentbisailon/fake-and-real-news-dataset>

Importing necessary libraries



```
[42]: import warnings
warnings.filterwarnings('ignore')
import numpy as np
import pandas as pd

import nltk
from nltk.sentiment import SentimentIntensityAnalyzer
import warnings
warnings.filterwarnings("ignore")
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

/kaggle/input/fake-and-real-news-dataset/True.csv
/kaggle/input/fake-and-real-news-dataset/Fake.csv
```

Loading the data

The screenshot shows a Google Colab environment with a Jupyter Notebook titled "NLP fake news detection". The notebook has three cells:

- Cell [43]:

```
true = pd.read_csv('/kaggle/input/fake-and-real-news-dataset/True.csv')
fake = pd.read_csv('/kaggle/input/fake-and-real-news-dataset/Fake.csv')
```
- Cell [44]:

```
fake['Category'] = 'fake'
fake
```
- Cell [44]: A preview of the data as a table.

	title	text	subject	date	Category
0	Donald Trump Sends Out Embarrassing New Year...	Donald Trump just couldn't wish all Americans ...	News	December 31, 2017	fake
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	fake
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	fake
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	fake
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	fake
...
23476	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016	fake
23477	ILLUSTIC? Yahoo Sattles E-mail Privacy (Iaccar...	21st Century Wire says It's a familiar theme...	Middle-east	January 16, 2016	fake

The right sidebar shows the "Notebook" panel with "Data" and "Models" sections. The "Models" section indicates "No models added" and provides a link to "Add a Kaggle model".

Browser tabs: (no subject) - lakshmikanan77, Fake and real news dataset, NLP fake news detection | Kaggle, Untitled11.ipynb - Colaboratory

URL: kaggle.com/code/lakshmikanan7781/nlp-fake-news-detection/edit

NLP fake news detection

File Edit View Run Add-ons Help

Run All Code

```
[44]: fake['Category'] = 'fake'
fake
```

	title	text	subject	date	Category
0	Donald Trump Sends Out Embarrassing New Year...	Donald Trump just couldn't wish all Americans ...	News	December 31, 2017	fake
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	fake
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	fake
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	fake
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	fake
...
23476	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016	fake
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It's a familiar theme ...	Middle-east	January 16, 2016	fake
23478	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016	fake
23479	How to Blow \$700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016	fake
23480	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016	fake

23481 rows x 5 columns

/kaggle/input/fake-and-real-news-d... (13h:6m)

Notebook: Data, Models (+ Add Models)

No models added. Add a Kaggle model.

Browser tabs: (no subject) - lakshmikanan77, Fake and real news dataset, NLP fake news detection | Kaggle, Untitled11.ipynb - Colaboratory

URL: kaggle.com/code/lakshmikanan7781/nlp-fake-news-detection/edit

NLP fake news detection

File Edit View Run Add-ons Help

Run All Code

```
true['Category'] = 'true'
true
```

	title	text	subject	date	Category
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017	true
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017	true
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017	true
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017	true
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017	true
...
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	true
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of L...	worldnews	August 22, 2017	true
21414	Minsk cultural hub becomes haven from authorities	MIANSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	true
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	true
21416	Indonesia to buy \$1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	true

21417 rows x 5 columns

/kaggle/input/fake-and-real-news-d... (13h:6m)

Notebook: Data, Models (+ Add Models)

No models added. Add a Kaggle model.

Browser tabs: (no subject) - lakshmikanan77, Fake and real news dataset, NLP fake news detection | Kaggle, Untitled11.ipynb - Colaboratory

Address bar: kaggle.com/code/lakshmikanan7781/nlp-fake-news-detection/edit

NLP fake news detection

File Edit View Run Add-ons Help

Code

```
data = pd.concat([fake, true], ignore_index = True)
data
```

[46]:

		title	text	subject	date	Category
0		Donald Trump Sends Out Embarrassing New Year...	Donald Trump just couldn't wish all Americans ...	News	December 31, 2017	fake
1		Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	fake
2		Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	fake
3		Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	fake
4		Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	fake
...	
44893		'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	true
44894		LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of L...	worldnews	August 22, 2017	true
44895		Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	true
44896		Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	true
44897		Indonesia to buy \$1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	true

44898 rows x 5 columns

Right sidebar: Notebook, Data, Models, Add Models, No models added, Add a Kaggle model

Browser tabs: (no subject) - lakshmikanan77, Fake and real news dataset, NLP fake news detection | Kaggle, Untitled11.ipynb - Colaboratory

Address bar: kaggle.com/code/lakshmikanan7781/nlp-fake-news-detection/edit

NLP fake news detection

File Edit View Run Add-ons Help

Code

```
data.shape
```

[47]: (44898, 5)

Right sidebar: Notebook, Data, Models, Add Models, No models added, Add a Kaggle model

Preprocessing

The image displays two sequential screenshots of a Jupyter Notebook titled "NLP fake news detection" on the Kaggle platform. The notebook is in a "Draft Session" and has been saved once.

Top Screenshot:

- Code Cell [48]:** `data['Category'].value_counts()`
- Output [48]:** A table showing the distribution of the 'Category' variable:

Category	count
false	23481
true	21417
- Code Cell [49]:** `from sklearn.preprocessing import LabelEncoder
encoder = LabelEncoder()
data['Category'] = encoder.fit_transform(data['Category'])`
- Code Cell [50]:** `data['Category']`
- Output [50]:** A small table showing the first five rows of the 'Category' column:

	Category
0	0
1	0
2	0
3	0
4	0

Bottom Screenshot:

- Code Cell [50]:** `data['Category']`
- Output [50]:** A table showing the first five rows of the 'Category' column, followed by a summary row:

	Category
0	0
1	0
2	0
3	0
4	0
44893	1
44894	1
44895	1
44896	1
44897	1

Name: Category, Length: 44898, dtype: int64
- Code Cell [51]:** `vectorizer = TfidfVectorizer()
title = vectorizer.fit_transform(data['title'])
title`
- Output [51]:** A message indicating the creation of a sparse matrix:

<44898x20896 sparse matrix of type '<class 'numpy.float64''>' with 546512 stored elements in Compressed Sparse Row format>

✓ Thus, the dataset has been loaded and preprocessed.