FAKE NEWS DETECTION

PRESENTED BY :
ABDELILAH EL FEDG

SUPERVISED BY :
PR. KHAMJANE AZIZ

OUTLINE :

- TOOLS USED
- GENERAL ARCHITECTURE
- DATA HANDLING
- MODEL BUILDING
- PRACTICAL PART

TOOLS USED

Development and Environment

Data Handling and Analysis

seaborn

matplotlib

pandas

Natural Language Processing

NLTK

spaCy

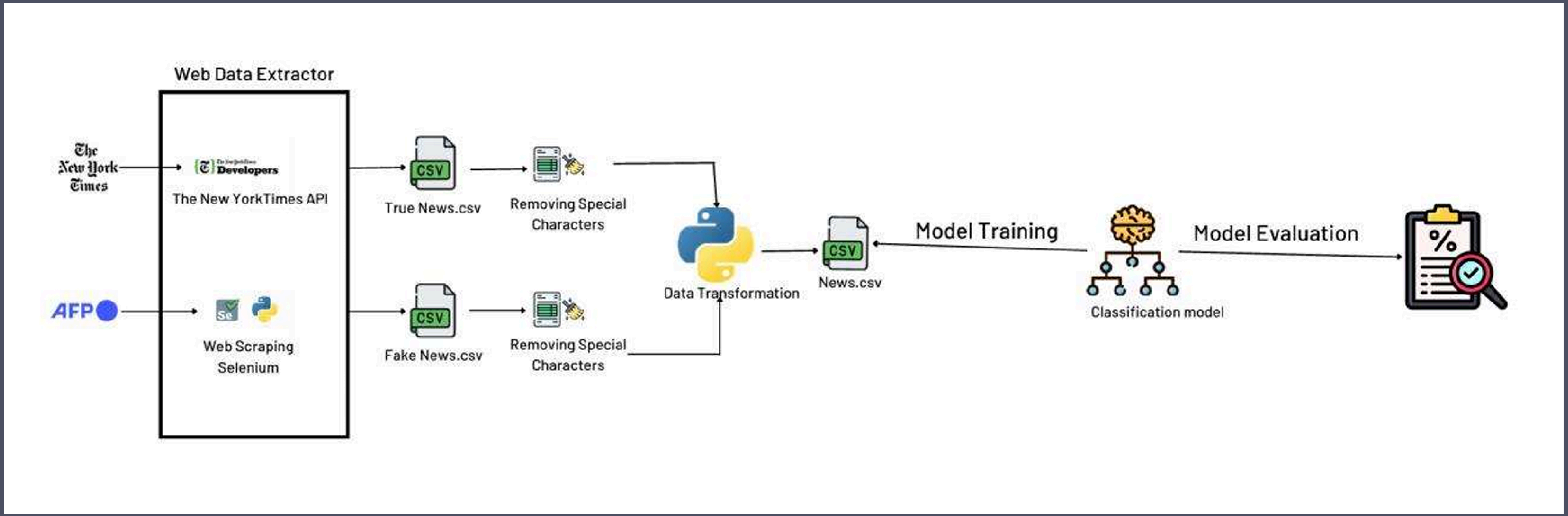
Image Processing and OCR

Tesseract

pillow

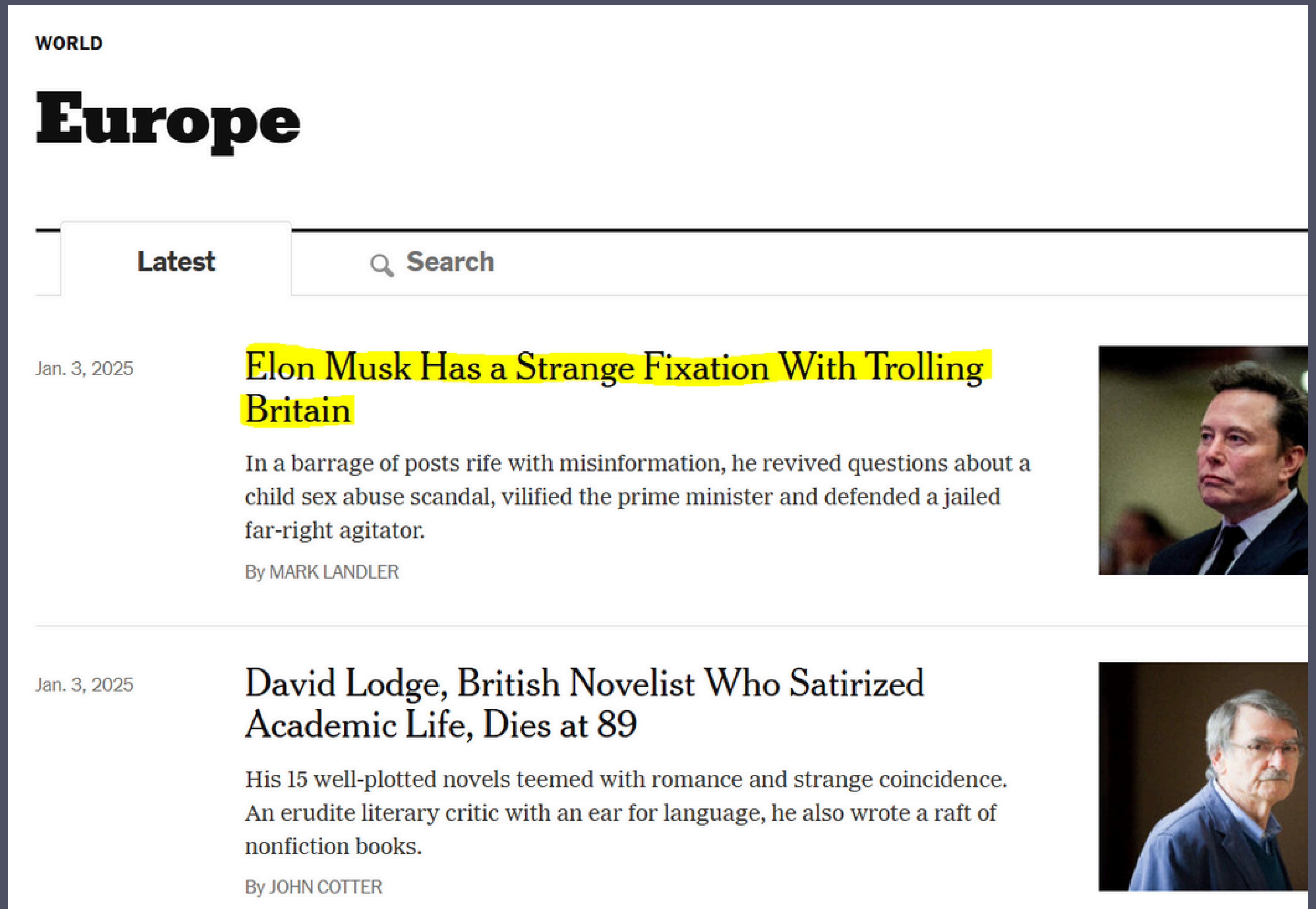
OpenCV

GENERAL ARCHITECTURE



DATA COLLECTION

The New York Times Website


WORLD

Europe

Latest | Search

Jan. 3, 2025

Elon Musk Has a Strange Fixation With Trolling Britain

In a barrage of posts rife with misinformation, he revived questions about a child sex abuse scandal, vilified the prime minister and defended a jailed far-right agitator.

By MARK LANDLER

Jan. 3, 2025

David Lodge, British Novelist Who Satirized Academic Life, Dies at 89

His 15 well-plotted novels teemed with romance and strange coincidence. An erudite literary critic with an ear for language, he also wrote a raft of nonfiction books.

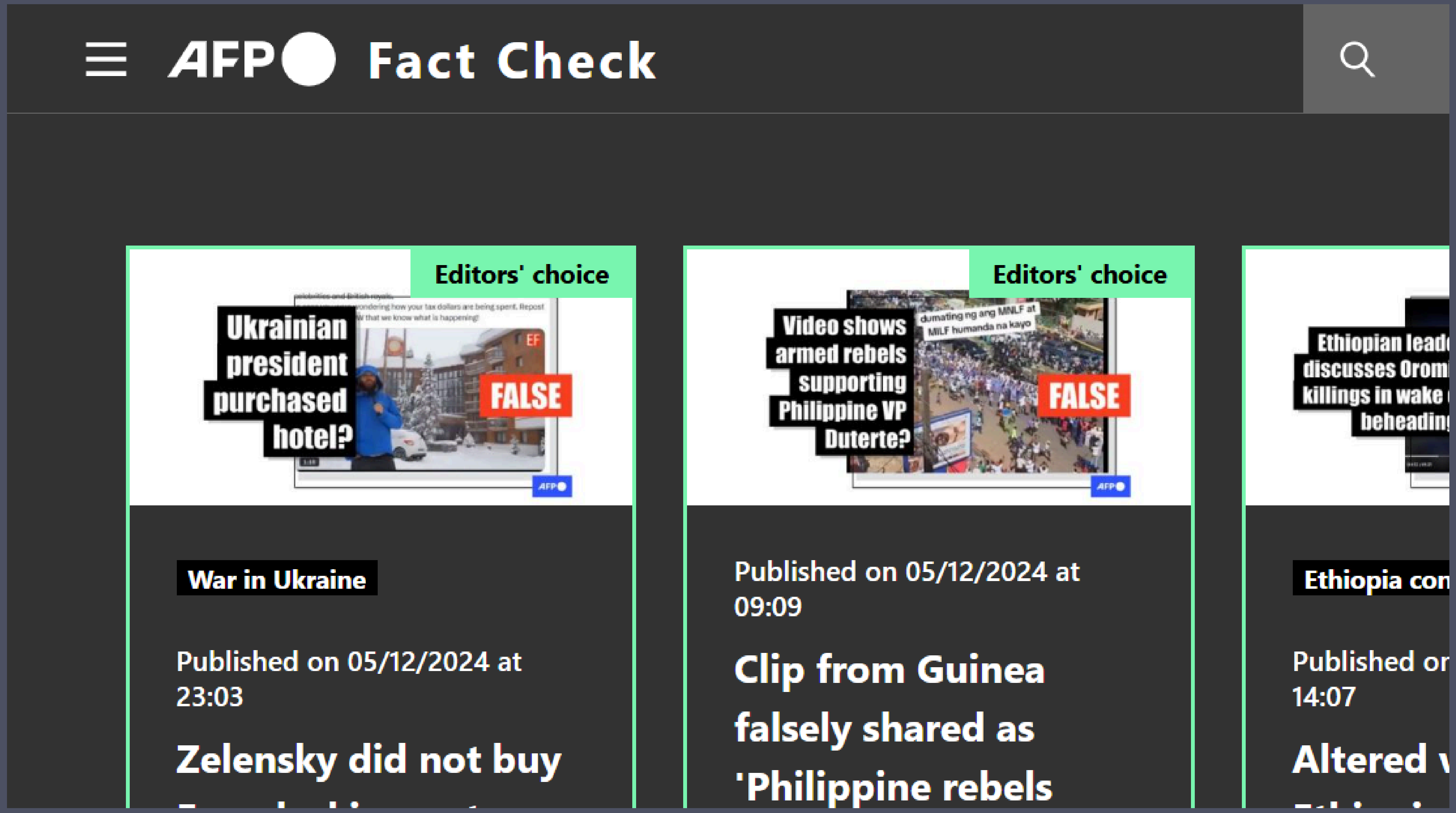
By JOHN COTTER

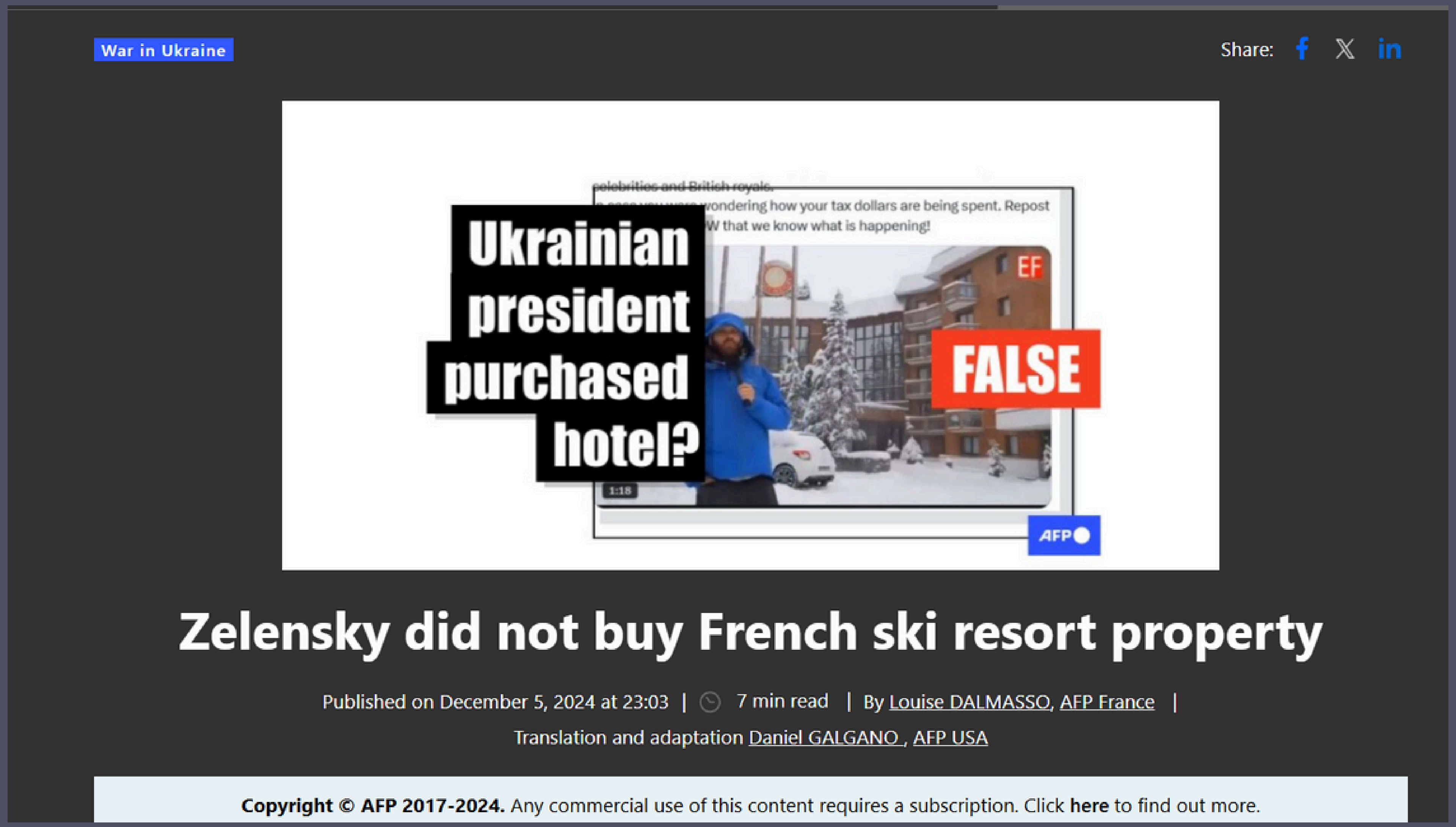
API response


```
{
  "response": {
    "meta": {
      "hits": 25,
      "time": 332,
      "offset": 0
    },
    "docs": [
      {
        "web_url": "http://thecaucus.blogs.nytimes.com/2012/01/01/virginia-at
        "snippet": "Virginia's attorney general on Sunday backed off of a pro
        "lead_paragraph": "DES MOINES -- Virginia's attorney general on Sunda
      }
    ],
    "facets": {
        "day_of_week": {
            "_type": "terms",
```

DATA COLLECTION

AFP fact check





Ukrainian President Volodymyr Zelensky has faced scrutiny for his offshore holdings as his country fights off Russia's invasion with Western economic assistance. But social media claims that he bought a luxury resort hotel in the French Alps are false; a Monaco-based company announced in 2023 that it had acquired the property and said the posts are inaccurate.

"Zelensky has purchased €88 million hotel in Courchevel ski resort in France. Courchevel is known to be wildly popular with billionaires, Hollywood celebrities and British royals," says a November 28, 2024 X post with tens of thousands of interactions.

"In case you were wondering how your tax dollars are being spent. Repost this So THEY KNOW that we know what is happening!"

The post includes a video of a man standing outside the resort supposedly purchased by the Ukrainian leader, speaking as if he were reporting for a television news broadcast. He claims Zelensky bought Le Palace des Neiges hotel at France's Courchevel ski resort through a Belize-based company called Film Heritage Inc.

DATA COLLECTION

Scraping

Fake Articles Infos


```
[
  {
    "title": "Altered video of old Ethiopian PM speech falsely linked to Oromia killings in 202
    "url": "https://factcheck.afp.com/doc.afp.com.36PN72Y",
    "publication_date": "Published on December 6, 2024 at 14:07",
    "content": [
      "“About Salale killings,” reads a text overlay in Amharic splashed across the top of a
      "Salale people – from a clan that belongs to Ethiopia’s Oromo – live in the north Shewa
      "The post, shared more than 4,400 times, shows Abiy in black clothes speaking to the ca
    ]
  },
  {
    "title": "Video shows Kenyan president challenging opposition party leader, not his former
    "url": "https://factcheck.afp.com/doc.afp.com.36PE4BL",
    "publication_date": "Published on December 5, 2024 at 16:33",
    "content": [
      "“‘If you have issues, let us meet in 2027’ President Ruto warns Rigathi Gachagua,” rea
      "",
```


Images Links


```
[
  {
    "https://factcheck.afp.com/doc.afp.com.36PN72Y": "https://factcheck.afp.com/sites/default/fi
  },
  {
    "https://factcheck.afp.com/doc.afp.com.36PE4BL": "https://factcheck.afp.com/sites/default/fi
  },
  {
    "https://factcheck.afp.com/doc.afp.com.36PH68T": "https://factcheck.afp.com/sites/default/fi
  },
  {
    "https://factcheck.afp.com/doc.afp.com.36P44LY": "https://factcheck.afp.com/sites/default/fi
  },
  {
    "https://factcheck.afp.com/doc.afp.com.36NK76B": "https://factcheck.afp.com/sites/default/fi
  },
```


Article’s title


```
[
  {
    "title": "Video shows ~ Fi\nPakistan police da\nheating women FALSE\n\nlining up for flour?
  },
  {
    "title": "Ukrainian\nprosecutor's = | =\nSOniN 5 Sear\nMonaco? \"+\nAFP@\n"
  },
  {
    "title": "Video shows -\nIndonesian B\nparty chairman q\nskydiving?! | |\nabe\n= AFP@®\n"
  },
  {
    "title": "First president to | x\nvisit restive Qo) a)\nprovince inthe 5 MISLEADING\nPhilip
  },
  {
    "title": "| a\n\nVideo shows Taylor\n\nSwift and Travis\nKelce falling off FALSE\n\nbeach s
  },
```

DATA PREPROCESSING

Removing Special Characters

True Articles :

```
import re

def clean_textTrueNews(text):
  if not isinstance(text, str):
      return ""
  text = re.sub(r'<[^>]+>', '', text)
  text = re.sub(r'\s+', ' ', text)
  text = re.sub(r'http\S+', '', text)
  text = re.sub(r'[^a-zA-Z0-9\s]', '', text)
  text = re.sub(r'@\w+', '', text)
  text = re.sub(r'#\w+', '', text)
  text = text.lower()
  return text.strip()
```

Before :

THE number of American visitors to South Africa dropped 13 percent in the months just after Sept. 11, the state tourism agency here says, but since then it has picked up and appears to be climbing. This is in contrast to countries like Kenya and Tanzania, where the bombings of two American embassies received considerable publicity four years ago. Travel agents are still reporting sharp declines in bookings in that region, as much as 25 percent in some cases. East Africa's tourist woes have meant more business for South Africa, which is increasingly viewed as a safe haven for people eager to glimpse the wildlife.

After :

the number of american visitors to south africa dropped 13 percent in the months just after sept 11 the state tourism agency here says but since then it has picked up and appears to be climbing this is in contrast to countries like kenya and tanzania where the bombings of two american embassies received considerable publicity four years ago travel agents are still reporting sharp declines in bookings in that region as much as 25 percent in some cases east africas tourist woes have meant more business for south africa which is increasingly viewed as a safe haven for people eager to glimpse the wildlife

DATA PREPROCESSING

Removing Special Characters

Fake Articles :

```python
def clean_textFakeNews(text):
    if not isinstance(text, str):
        return ""

    text = re.sub(r"\[\"|\"\]", "", text)
    # Étape 1 : Extraction des textes entre guillemets
    quotes = re.findall(r'"(.*?)"|“(.*?)”', text)
    if quotes:
        extracted_texts = [''.join(q) for q in quotes]
        text = ' '.join(extracted_texts).strip()
    else:
        return ""

    # Étape 2 : Suppressions progressives
    text = re.sub(r'<[^>]+>', '', text)  # Supprimer balises HTML
    text = re.sub(r'http\S+', '', text)  # Supprimer URLs
    text = re.sub(r'@\w+', '', text)     # Supprimer mentions
    text = re.sub(r'#\w+', '', text)     # Supprimer hashtags
    text = re.sub(r'[^a-zA-Z0-9\s]', '', text)  # Conserver lettres,
    text = re.sub(r'\s+', ' ', text)  # Réduire espaces multiples
    text = text.lower()  # Tout en minuscule

    return text.strip()
```

Before :

content
['"Walz changed his states (sic) flag to look exactly like the Somali flag, to pander to the Somali residents of his state," says an X post published August 6, 2024, shortly after Vice President Kamala Harris selected the 60-year-old governor as her running mate in the presidential race.', 'Another X user wrote: "He turned his home state into Somalia and now he wants to do the same to our country."', 'The claims were amplified by X owner and Tesla founder Elon Musk, podcast host Joe Rogan and Fox News host Jesse Watters. Similar allegations shot across TikTok, Facebook and Instagram.']

After :

content
walz changed his states sic flag to look exactly like the somali flag to pander to the somali residents of his state he turned his home state into somalia and now he wants to do the same to our country

DATA PREPROCESSING

Removing Special Characters

Fake Articles (Title) :

```python
def clean_titleFakeNews(text):
  custom_words = set(["false", "misleading", "altered"])
  if not isinstance(text, str):  # Vérifie si la valeur est une chaîne
    return ""
  text = re.sub(r'[^a-zA-Z\s]', '', text)
  text = re.sub(r'\s+', ' ', text).strip()
  text = text.lower()
  text = ' '.join([word for word in text.split() if len(word) > 3])
  text = ' '.join([word for word in text.split() if word.lower() not in custom_words])

  lettres_repete = r'\b\w*([a-zA-Z])\1{2,}\w*\b'
  text = re.sub(lettres_repete, '', text)
  text = re.sub(r'\s+', ' ', text).strip()

  return text
```

Before :

title
Video shows ~ Fi Pakistan police da heating women FALSE lining up for flour? a a er AFP@
Ukrainian prosecutor's = | = SOniN 5 Sear Monaco? "+ AFP@
Video shows - Indonesian B party chairman q skydiving?! | | abe = AFP@®
First president to | x visit restive Qo) a) province inthe 5 MISLEADING Philippines? AFP@®
| a Video shows Taylor Swift and Travis Kelce falling off FALSE beach swing?

After :

video shows pakistan police heating women lining flour
ukrainian prosecutors sonin sear monaco
video shows indonesian party chairman skydiving
first president visit restive province inthe philippines
video shows taylor swift travis kelce falling beach swing

DATA PREPROCESSING

Removing Stop Words

Stop Words include :

- a
- I
- the
- in
- of
- for
- at
- to
- on
- with
- from

Example :

With Stop Words :	Without Stop Words :
this article is also a weekly newsletter sign up for racerelated here	article also weekly newsletter sign racerelated

DATA PREPROCESSING

Tokenization and Lemmatization

Use of “en_core_web_sm”

```
import spacy

nlp = spacy.load("en_core_web_sm")

def lemmatize_text(text):
    # Traitez le texte avec spaCy, cela effectue la tokenisation et la lemmatisation
    doc = nlp(text)
    # Retourne la chaîne lemmatisée en un seul passage
    return " ".join([token.lemma_ for token in doc])


print(News['label'].value_counts())
```

Before	After
article also weekly newsletter sign racerelated	['article', 'also', 'weekly', 'newsletter', 'sign', 'racerelated']

DATA PREPROCESSING

TF-IDF

$$TF(t,d) = \frac{number\ of\ times\ t\ appears\ in\ d}{total\ number\ of\ terms\ in\ d}$$

$$IDF(t) = log\frac{N}{1+df}$$

$$TF-IDF(t,d) = TF(t,d) * IDF(t)$$

After application of TF-IDF :

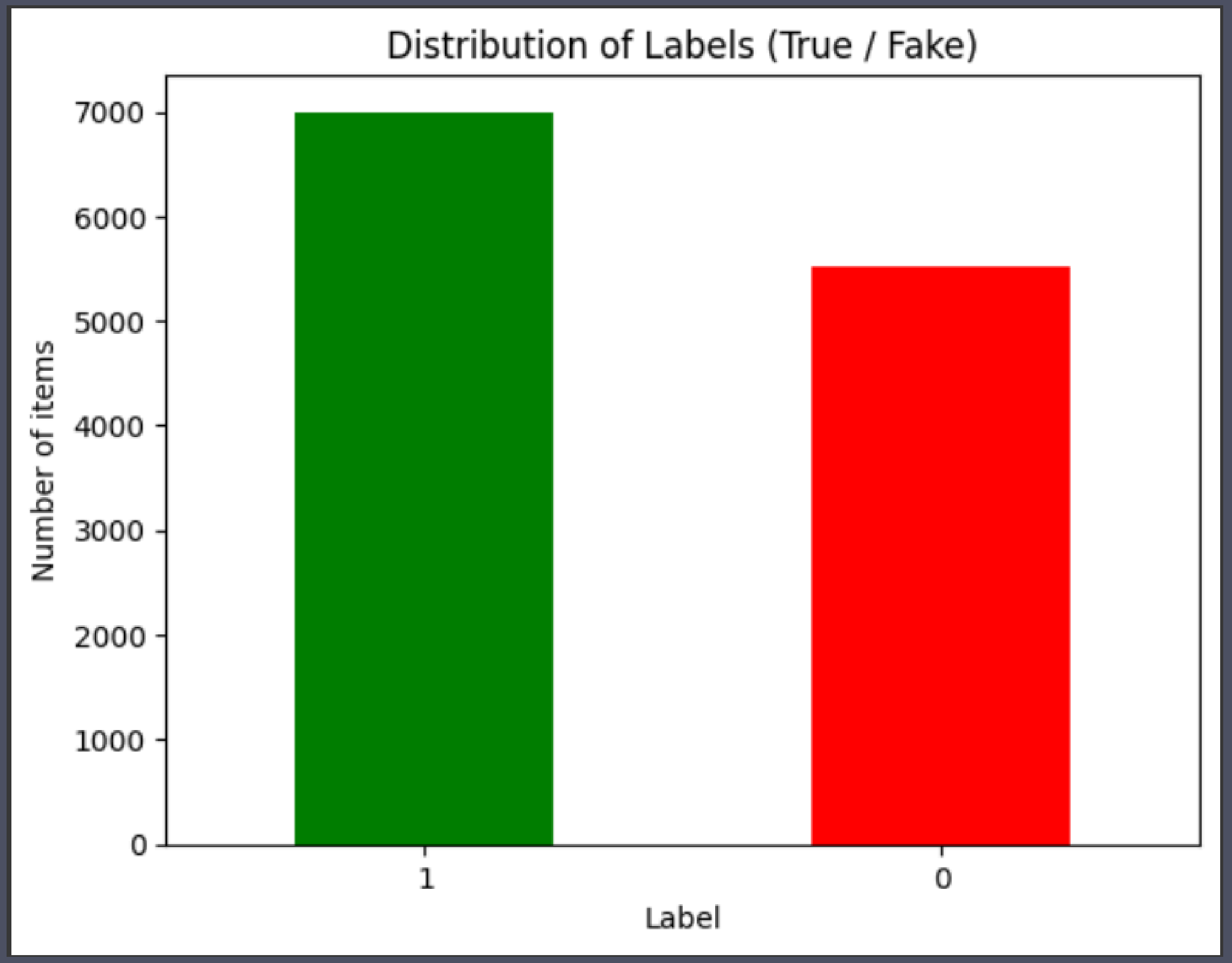
						...	zurmi	zyad	zyakhala	zygmunt	zyl
0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0

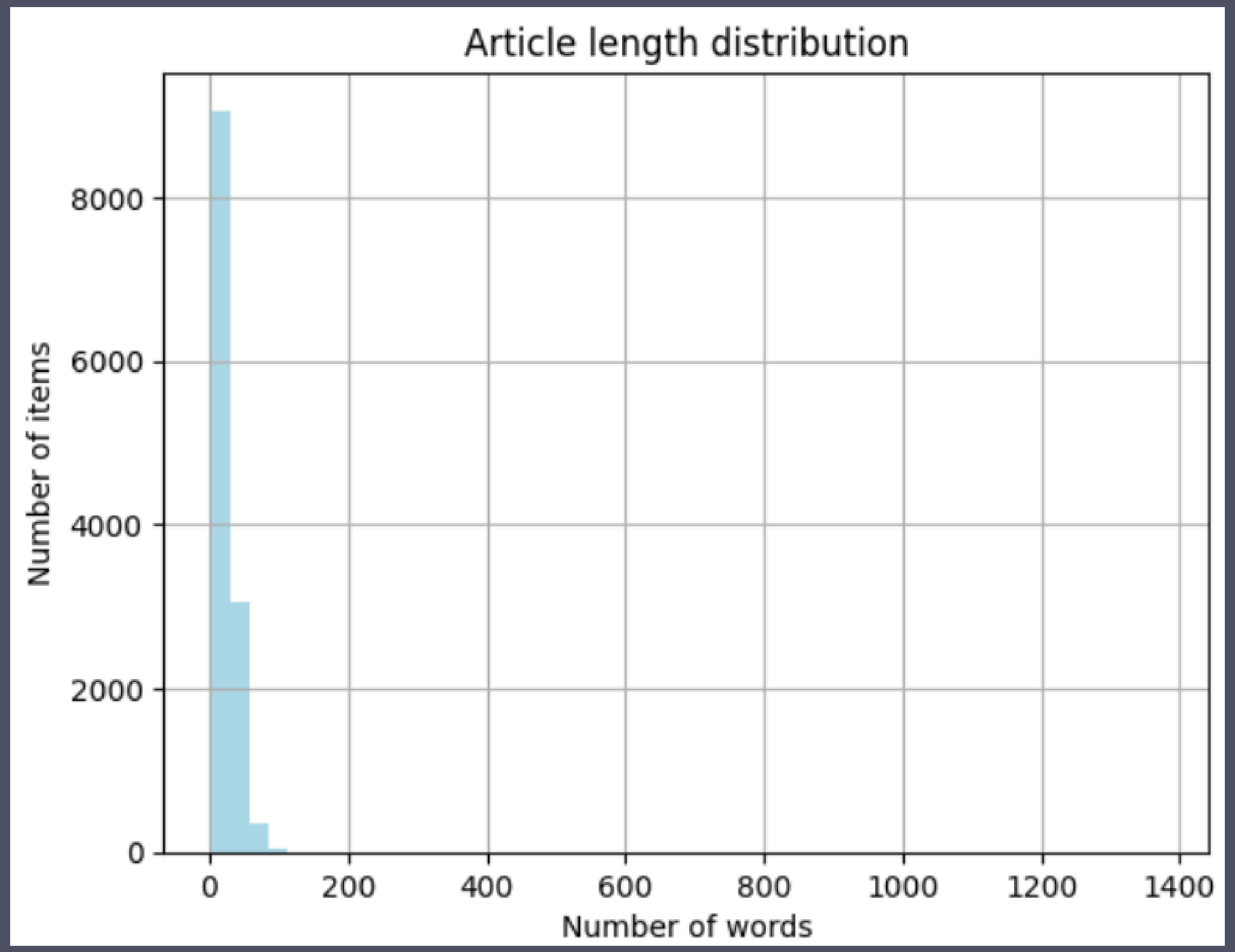
[10 rows x 33176 columns]

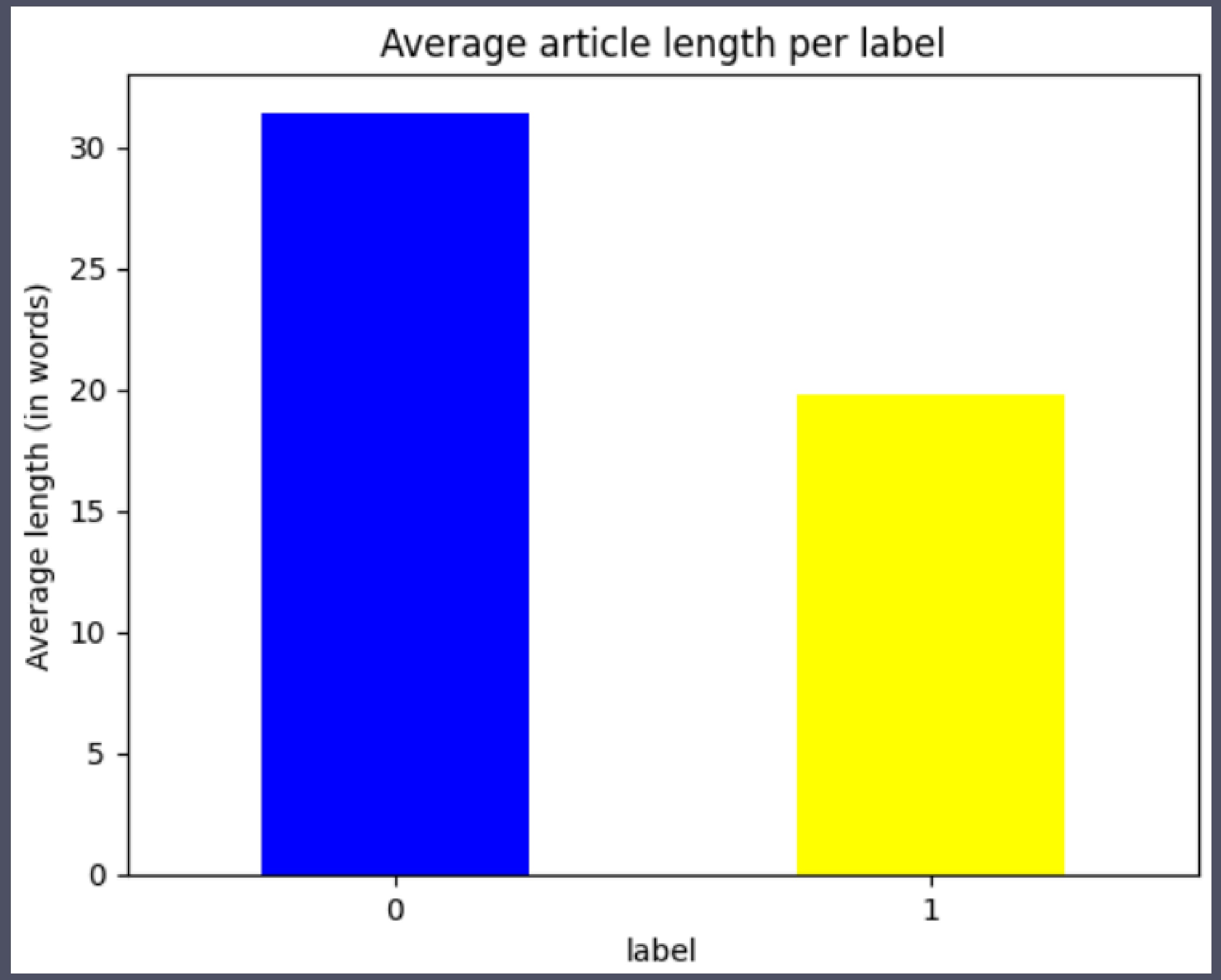
DATA VISUALIZATION

True Articles : 5524

Fake Articles : 6996







MODEL BUILDING

Logistic Regression :

Hyperparameter	Values
C	0.01, 0.1, 1
Penalty	l2, l1
Solver	liblinear, saga

Best hyperparameters :

C = 1 Solver = saga

Penalty = l2

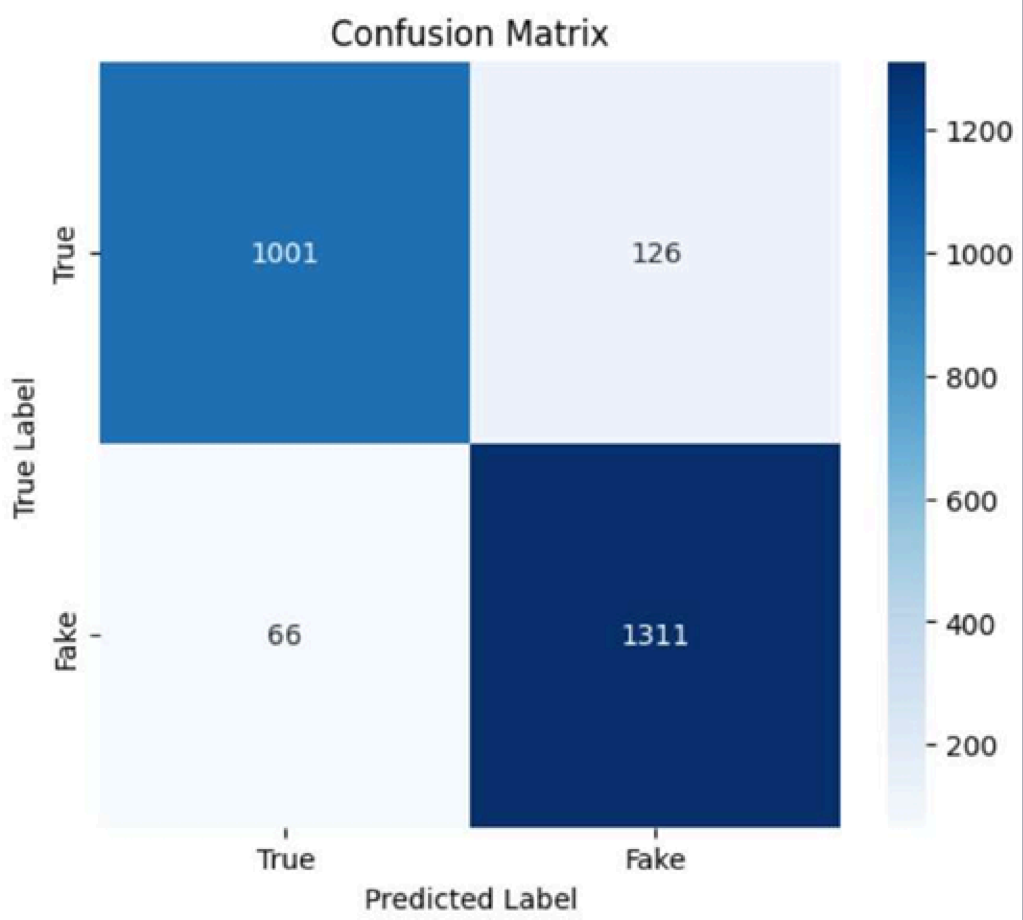
Classification Report :

	precision	recall	f1-score	support
0	0.94	0.89	0.91	1127
1	0.91	0.95	0.93	1377
accuracy			0.92	2504
macro avg	0.93	0.92	0.92	2504
weighted avg	0.92	0.92	0.92	2504

Accuracy for test data : 0.923

Accuracy for train data : 0.969

Confusion Matrix :



MODEL BUILDING

Bayesian classification :

Hyperparameter	Values
Alpha	0.01, 0.5, 1, 2, 5

Classification Report :

	precision	recall	f1-score	support
0	0.92	0.90	0.91	1127
1	0.92	0.94	0.93	1377
accuracy			0.92	2504
macro avg	0.92	0.92	0.92	2504
weighted avg	0.92	0.92	0.92	2504

Best hyperparameters :

Alpha = 0.5

Accuracy for test data : 0.922

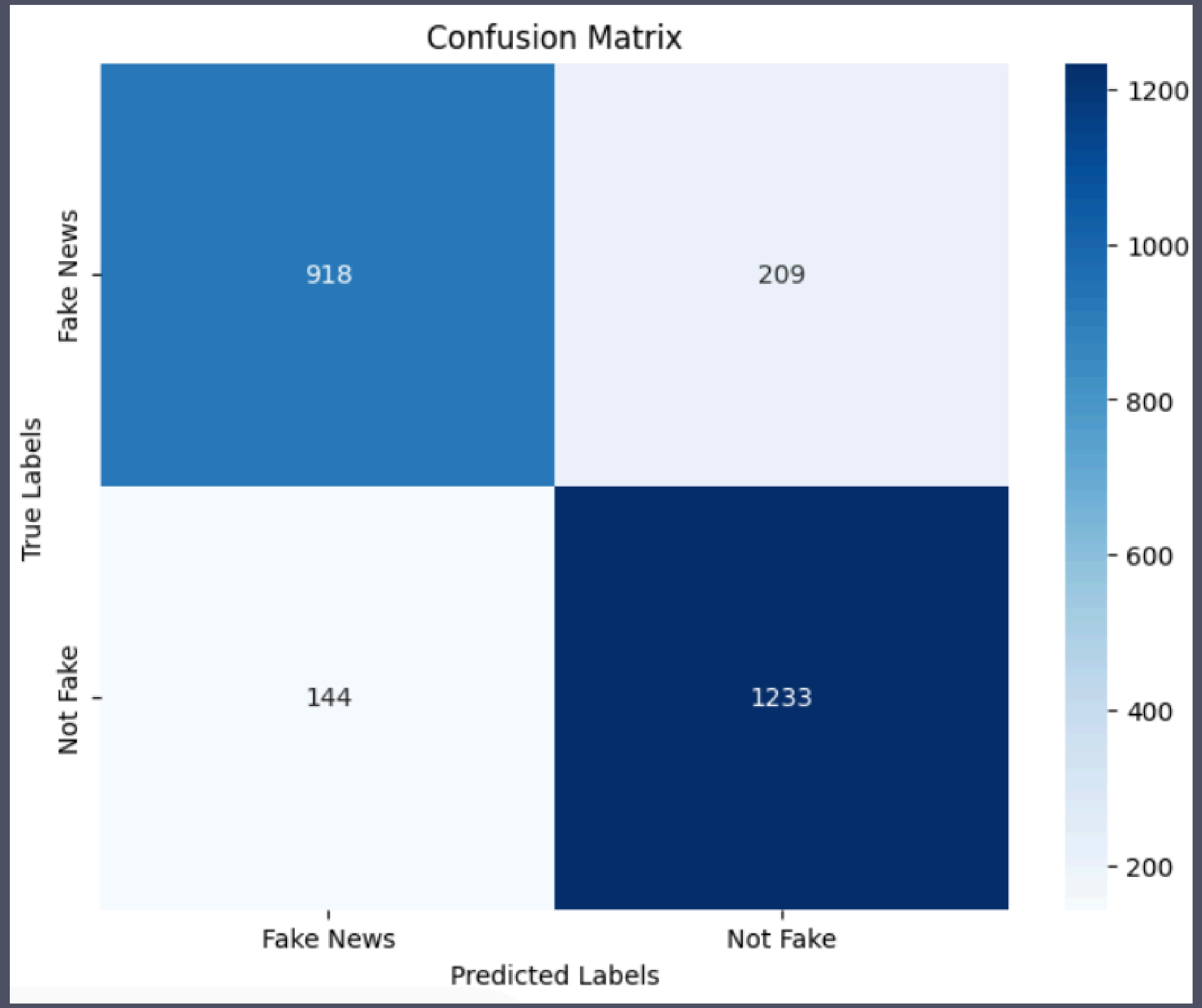
Accuracy for train data : 0.979

MODEL BUILDING

Decision Tree :

Hyperparameter	Values
max_depth	range(150. 200)
min_samples_split	200, 300
min_samples_leaf	5, 10
max_leaf_nodes	200, 300
min_impurity_decrease	0.0001, 0.001, 0.01

Confusion Matrix :



Accuracy for test data : 0.85

Accuracy for train data : 0.87

PRACTICAL PART