

Detecting Fake vs Real News using Machine Learning



Software Engineering Project: EEC 626
User Manual

Anubhuti Dayal – 2824826

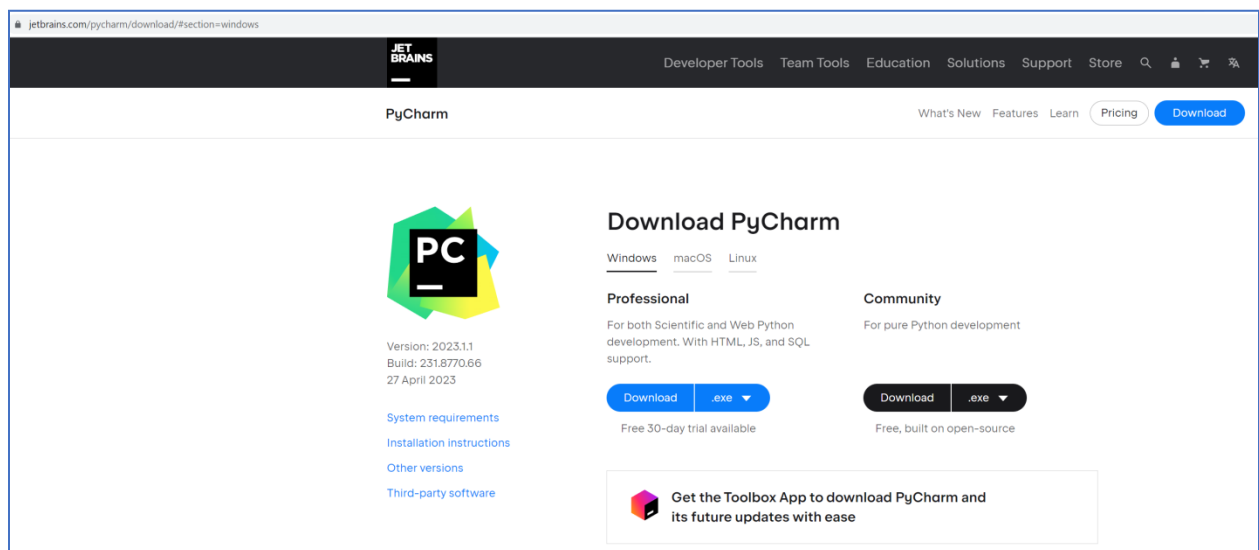
Oyindoubra Timi – 2822556

User Manual

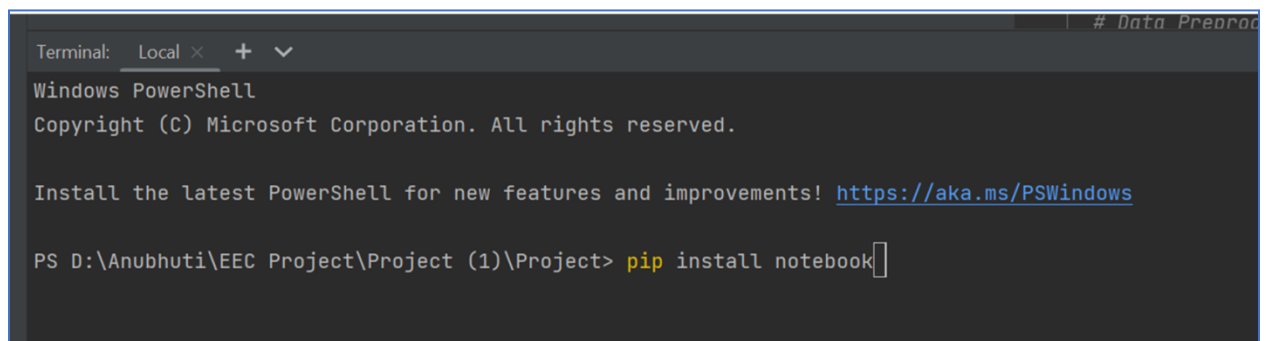
This is user manual for the project code “Detecting Fake vs Real News using Machine Learning”. Please follow below steps to run the program in your machine

1. Download and install PyCharm or any python supported IDE. For this project, PyCharm and VS Code were used.

You can follow this link - https://youtu.be/qC6-Uv9m_Ls
<https://www.jetbrains.com/pycharm/download/#section=windows>



2. Open the project folder in the PyCharm environment on your machine after installation
3. Open a terminal on the IDE
4. If jupyterlab is not already installed in your system, type “pip install notebook” and hit “Enter” on your keyboard



- Next, type “jupyter notebook” and hit “Enter.” This should open a tab on your browser that shows all the documents and files in the project folder
- Scroll to “FakeNews.ipynb” file and click on it. This should open in a new tab

```

Settings  Help
File Edit View Insert Cell Help
[*]: import re
import pandas as pd
import numpy as np
import random
from wordcloud import WordCloud
from tqdm import tqdm
import matplotlib.pyplot as plt
from collections import Counter
import seaborn as sns
import torch
from transformers import BertForSequenceClassification, BertTokenizer
from transformers.file_utils import is_tf_available, is_torch_tpu_available, is_torch_available
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
import nltk
from nltk.stem.porter import PorterStemmer
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
nltk.download('stopwords')
nltk.download('punkt')
nltk.download('wordnet')

[*]: train_data = pd.read_csv('new dataset.csv')
train_data = train_data.loc[:, ~train_data.columns.str.contains('^Unnamed')]
train_data

[*]: # Data Preprocessing
def preprocessing_data(text):
    text = str(text).replace('http[\\w/\\.\\-\\+]', ' ')
    words = re.sub('[^\\w\\s]', '', text).split()
    text = ' '.join([nltk.stem.WordNetLemmatizer().lemmatize(word) for word in words if word not in stopwords.words('english')])
    return text

[*]: train_data['text'] = train_data.text.apply(preprocessing_data)

[*]: sns.countplot(data = train_data, x = 'class')

[*]: realCloud = ' '.join(train_data[train_data['class'] == 1]['text'])
words_cloud = WordCloud(background_color='black', min_font_size = 10, max_font_size = 100, include_numbers = False, collocations=False, width=2000, height=750)
plt.figure(figsize=(15, 30))
plt.imshow(words_cloud.generate(realCloud))
plt.axis('off')
plt.show()

[*]: fakeCloud = ' '.join(train_data[train_data['class'] == 0]['text'])
words_cloud = WordCloud(background_color='black', min_font_size = 10, max_font_size = 100, include_numbers = False, collocations=False, width=2000, height=750)
plt.figure(figsize=(15, 30))
plt.imshow(words_cloud.generate(fakeCloud))
plt.axis('off')
plt.show()

```

- Run each cell. Some cells might take hours or days to finish running depending on the speed of your device
- After running the last cell, an input box will be displayed. **Input your news in the box and hit "Enter"**
- The result will be displayed to tell you if the news is real or fake

```
def get_prediction(text, convert_to_label=False):
    inputs = bert_token(text, padding=True, truncation=True, max_length=512, return_tensors="pt")
    outputs = new_train_model(**inputs)
    probs = outputs[0].softmax(1)
    d = {0: "Fake", 1: "Real"}
    if convert_to_label:
        return d[int(probs.argmax())]
    else:
        return int(probs.argmax())

news = str(input())

get_prediction(news, convert_to_label=True)
```

Former Vice President Mike Pence testified on Thursday to a federal grand jury investigating the aftermath of the 2020 election and the actions of then-President Donald Trump and others, sources familiar with the matter told CNN. The testimony marks a momentous juncture in the criminal investigation and the first time in modern history a vice president has been compelled to testify about the president he served beside. Pence testified for more than five hours, a source familiar with the matter told CNN, and while adviser Marc Short did not confirm the appearance on Thursday, he addressed the legal back-and-forth over the testimony. "I think that the vice president, you know, had his own case based on the Speech and Debate Clause. He was pleased that for the first time a judge acknowledged that it applied to the vice president of the United States," Short said in an interview on NewsNation afterward. "But he was willing to comply with the law, and courts have ordered him to testify."

'Real'