

snm

July 13, 2025

```
[1]: !pip install numpy pandas matplotlib seaborn scikit-learn
```

```
Requirement already satisfied: numpy in c:\users\naman\anaconda3\envs\ml-dl-  
env\lib\site-packages (2.1.3)  
Requirement already satisfied: pandas in c:\users\naman\anaconda3\envs\ml-dl-  
env\lib\site-packages (2.3.1)  
Requirement already satisfied: matplotlib in c:\users\naman\anaconda3\envs\ml-  
dl-env\lib\site-packages (3.10.3)  
Requirement already satisfied: seaborn in c:\users\naman\anaconda3\envs\ml-dl-  
env\lib\site-packages (0.13.2)  
Requirement already satisfied: scikit-learn in c:\users\naman\anaconda3\envs\ml-  
dl-env\lib\site-packages (1.6.1)  
Requirement already satisfied: python-dateutil>=2.8.2 in  
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from pandas)  
(2.9.0.post0)  
Requirement already satisfied: pytz>=2020.1 in c:\users\naman\anaconda3\envs\ml-  
dl-env\lib\site-packages (from pandas) (2025.2)  
Requirement already satisfied: tzdata>=2022.7 in  
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from pandas) (2025.2)  
Requirement already satisfied: contourpy>=1.0.1 in  
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from matplotlib)  
(1.3.2)  
Requirement already satisfied: cycler>=0.10 in c:\users\naman\anaconda3\envs\ml-  
dl-env\lib\site-packages (from matplotlib) (0.12.1)  
Requirement already satisfied: fonttools>=4.22.0 in  
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from matplotlib)  
(4.58.5)  
Requirement already satisfied: kiwisolver>=1.3.1 in  
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from matplotlib)  
(1.4.8)  
Requirement already satisfied: packaging>=20.0 in  
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from matplotlib)  
(25.0)  
Requirement already satisfied: pillow>=8 in c:\users\naman\anaconda3\envs\ml-dl-  
env\lib\site-packages (from matplotlib) (11.3.0)  
Requirement already satisfied: pyparsing>=2.3.1 in  
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from matplotlib)  
(3.2.3)
```

Requirement already satisfied: scipy>=1.6.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from scikit-learn) (1.15.3)

Requirement already satisfied: joblib>=1.2.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from scikit-learn) (1.5.1)

Requirement already satisfied: threadpoolctl>=3.1.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from scikit-learn) (3.6.0)

Requirement already satisfied: six>=1.5 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)

[2]: !pip install datasets

Requirement already satisfied: datasets in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (3.6.0)

Requirement already satisfied: filelock in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from datasets) (3.18.0)

Requirement already satisfied: numpy>=1.17 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from datasets) (2.1.3)

Requirement already satisfied: pyarrow>=15.0.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from datasets) (20.0.0)

Requirement already satisfied: dill<0.3.9,>=0.3.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from datasets) (0.3.8)

Requirement already satisfied: pandas in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from datasets) (2.3.1)

Requirement already satisfied: requests>=2.32.2 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from datasets) (2.32.4)

Requirement already satisfied: tqdm>=4.66.3 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from datasets) (4.67.1)

Requirement already satisfied: xxhash in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from datasets) (3.5.0)

Requirement already satisfied: multiprocessing in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from datasets) (0.70.16)

Requirement already satisfied: fsspec<=2025.3.0,>=2023.1.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (2025.3.0)

Requirement already satisfied: huggingface-hub>=0.24.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from datasets) (0.33.2)

Requirement already satisfied: packaging in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from datasets) (25.0)

Requirement already satisfied: pyyaml>=5.1 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from datasets) (6.0.2)

Requirement already satisfied: aiohttp!=4.0.0a0,!4.0.0a1 in

```

c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (3.12.13)
Requirement already satisfied: aiohappyeyeballs>=2.5.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (2.6.1)
Requirement already satisfied: aiosignal>=1.1.2 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.4.0)
Requirement already satisfied: async-timeout<6.0,>=4.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (5.0.1)
Requirement already satisfied: attrs>=17.3.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets)
(25.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.7.0)
Requirement already satisfied: multidict<7.0,>=4.5 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (6.6.3)
Requirement already satisfied: propcache>=0.2.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (0.3.2)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets)
(1.20.1)
Requirement already satisfied: typing-extensions>=4.1.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from multidict<7.0,>=
4.5->aiohttp!=4.0.0a0,!4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets)
(4.14.1)
Requirement already satisfied: idna>=2.0 in c:\users\naman\anaconda3\envs\ml-dl-
env\lib\site-packages (from yarl<2.0,>=1.17.0->aiohttp!=4.0.0a0,!4.0.0a1->fsspe
c[http]<=2025.3.0,>=2023.1.0->datasets) (3.10)
Requirement already satisfied: charset_normalizer<4,>=2 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
requests>=2.32.2->datasets) (3.4.2)
Requirement already satisfied: urllib3<3,>=1.21.1 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
requests>=2.32.2->datasets) (2.5.0)
Requirement already satisfied: certifi>=2017.4.17 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
requests>=2.32.2->datasets) (2025.6.15)
Requirement already satisfied: colorama in c:\users\naman\anaconda3\envs\ml-dl-
env\lib\site-packages (from tqdm>=4.66.3->datasets) (0.4.6)
Requirement already satisfied: python-dateutil>=2.8.2 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from

```

```
pandas->datasets) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\naman\anaconda3\envs\ml-
dl-env\lib\site-packages (from pandas->datasets) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
pandas->datasets) (2025.2)
Requirement already satisfied: six>=1.5 in c:\users\naman\anaconda3\envs\ml-dl-
env\lib\site-packages (from python-dateutil>=2.8.2->pandas->datasets) (1.17.0)
```

```
[3]: from datasets import load_dataset

ds = load_dataset("zeroshot/twitter-financial-news-sentiment")
```

```
c:\Users\naman\anaconda3\envs\ml-dl-env\lib\site-packages\tqdm\auto.py:21:
TqdmWarning: IProgress not found. Please update jupyter and ipywidgets. See
https://ipywidgets.readthedocs.io/en/stable/user_install.html
    from .autonotebook import tqdm as notebook_tqdm
```

```
[4]: import pandas as pd

print(ds)
```

```
DatasetDict({
  train: Dataset({
    features: ['text', 'label'],
    num_rows: 9543
  })
  validation: Dataset({
    features: ['text', 'label'],
    num_rows: 2388
  })
})
```

```
[5]: train_ds = ds['train']
```

```
[6]: print(train_ds[0])
```

```
{'text': '$BYND - JPMorgan reels in expectations on Beyond Meat
https://t.co/bd0xbFGjkT', 'label': 0}
```

```
[7]: train_ds
```

```
[7]: Dataset({
  features: ['text', 'label'],
  num_rows: 9543
})
```

```
[8]: !pip install spacy
```

Requirement already satisfied: spacy in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (3.8.7)

Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (3.0.12)

Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (1.0.5)

Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (1.0.13)

Requirement already satisfied: cymem<2.1.0,>=2.0.2 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (2.0.11)

Requirement already satisfied: preshed<3.1.0,>=3.0.2 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (3.0.10)

Requirement already satisfied: thinc<8.4.0,>=8.3.4 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (8.3.6)

Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (1.1.3)

Requirement already satisfied: srsly<3.0.0,>=2.4.3 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (2.5.1)

Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (2.0.10)

Requirement already satisfied: weasel<0.5.0,>=0.1.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (0.4.1)

Requirement already satisfied: typer<1.0.0,>=0.3.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (0.16.0)

Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (4.67.1)

Requirement already satisfied: numpy>=1.19.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (2.1.3)

Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (2.32.4)

Requirement already satisfied: pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (2.11.7)

Requirement already satisfied: jinja2 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (3.1.6)

Requirement already satisfied: setuptools in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (78.1.1)

Requirement already satisfied: packaging>=20.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (25.0)

Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from spacy) (3.5.0)

Requirement already satisfied: language-data>=1.2 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from langcodes<4.0.0,>=3.2.0->spacy) (1.3.0)

Requirement already satisfied: annotated-types>=0.6.0 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (0.7.0)

Requirement already satisfied: pydantic-core==2.33.2 in c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from

pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (2.33.2)
Requirement already satisfied: typing-extensions>=4.12.2 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (4.14.1)
Requirement already satisfied: typing-inspection>=0.4.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (0.4.1)
Requirement already satisfied: charset_normalizer<4,>=2 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
requests<3.0.0,>=2.13.0->spacy) (3.4.2)
Requirement already satisfied: idna<4,>=2.5 in c:\users\naman\anaconda3\envs\ml-
dl-env\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
requests<3.0.0,>=2.13.0->spacy) (2.5.0)
Requirement already satisfied: certifi>=2017.4.17 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
requests<3.0.0,>=2.13.0->spacy) (2025.6.15)
Requirement already satisfied: blis<1.4.0,>=1.3.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
thinc<8.4.0,>=8.3.4->spacy) (1.3.0)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
thinc<8.4.0,>=8.3.4->spacy) (0.1.5)
Requirement already satisfied: colorama in c:\users\naman\anaconda3\envs\ml-dl-
env\lib\site-packages (from tqdm<5.0.0,>=4.38.0->spacy) (0.4.6)
Requirement already satisfied: click>=8.0.0 in c:\users\naman\anaconda3\envs\ml-
dl-env\lib\site-packages (from typer<1.0.0,>=0.3.0->spacy) (8.2.1)
Requirement already satisfied: shellingham>=1.3.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
typer<1.0.0,>=0.3.0->spacy) (1.5.4)
Requirement already satisfied: rich>=10.11.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
typer<1.0.0,>=0.3.0->spacy) (14.0.0)
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
weasel<0.5.0,>=0.1.0->spacy) (0.21.1)
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
weasel<0.5.0,>=0.1.0->spacy) (7.3.0.post1)
Requirement already satisfied: wrapt in c:\users\naman\anaconda3\envs\ml-dl-
env\lib\site-packages (from smart-
open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.1.0->spacy) (1.17.2)
Requirement already satisfied: marisa-trie>=1.1.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from language-
data>=1.2->langcodes<4.0.0,>=3.2.0->spacy) (1.2.1)
Requirement already satisfied: markdown-it-py>=2.2.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from

```
rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from
rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (2.19.2)
Requirement already satisfied: mdurl~=0.1 in c:\users\naman\anaconda3\envs\ml-
dl-env\lib\site-packages (from markdown-it-
py>=2.2.0->rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (0.1.2)
Requirement already satisfied: MarkupSafe>=2.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from jinja2->spacy)
(3.0.2)
```

```
[9]: import numpy as np
import spacy
```

```
[10]: train_ds = ds['train']
train_df = pd.DataFrame(train_ds)
train_df.to_csv("train1.csv", index=False)
```

```
[11]: df = pd.read_csv("train1.csv")
```

```
[12]: !pip install imbalanced-learn
```

```
Requirement already satisfied: imbalanced-learn in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (0.13.0)
Requirement already satisfied: numpy<3,>=1.24.3 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from imbalanced-
learn) (2.1.3)
Requirement already satisfied: scipy<2,>=1.10.1 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from imbalanced-
learn) (1.15.3)
Requirement already satisfied: scikit-learn<2,>=1.3.2 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from imbalanced-
learn) (1.6.1)
Requirement already satisfied: sklearn-compat<1,>=0.1 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from imbalanced-
learn) (0.1.3)
Requirement already satisfied: joblib<2,>=1.1.1 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from imbalanced-
learn) (1.5.1)
Requirement already satisfied: threadpoolctl<4,>=2.0.0 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from imbalanced-
learn) (3.6.0)
```

```
[13]: import re
```

```
[14]: def clean(text):
cleaned = re.compile(r'<.*?>')
```

```

    return re.sub(cleaned, '', text)

df.text = df.text.apply(clean)

```

```

[15]: #Remove special characters
def is_special(text):
    rem = ''
    for i in text:
        if i.isalnum():
            rem = rem + i
        else:
            rem = rem + ' '
    return rem

df.text = df.text.apply(is_special)

```

```

[16]: # # Remove URLs
# def remove_urls(text):
#     url_pattern = r'https?:\/\/\S+\/www\.\S+'
#     return re.sub(url_pattern, '', text)

# # Apply to your DataFrame
# df['text'] = df['text'].apply(remove_urls)

```

```

[17]: #Convert everything to lowercase
def to_lower(text):
    return text.lower()

df.text = df.text.apply(to_lower)

```

```

[18]: !pip install nltk

```

```

Requirement already satisfied: nltk in c:\users\naman\anaconda3\envs\ml-dl-
env\lib\site-packages (3.9.1)
Requirement already satisfied: click in c:\users\naman\anaconda3\envs\ml-dl-
env\lib\site-packages (from nltk) (8.2.1)
Requirement already satisfied: joblib in c:\users\naman\anaconda3\envs\ml-dl-
env\lib\site-packages (from nltk) (1.5.1)
Requirement already satisfied: regex<=2021.8.3 in
c:\users\naman\anaconda3\envs\ml-dl-env\lib\site-packages (from nltk)
(2024.11.6)
Requirement already satisfied: tqdm in c:\users\naman\anaconda3\envs\ml-dl-
env\lib\site-packages (from nltk) (4.67.1)
Requirement already satisfied: colorama in c:\users\naman\anaconda3\envs\ml-dl-
env\lib\site-packages (from click->nltk) (0.4.6)

```



```
[19]: from nltk.corpus import stopwords
      from nltk.tokenize import word_tokenize
```

```
[20]: #Remove stopwords
      def rem_stopwords(text):
          stop_words = set(stopwords.words('english'))
          words = word_tokenize(text)
          return [w for w in words if w not in stop_words]

      df.text = df.text.apply(rem_stopwords)
```

```
[44]: import nltk
      from nltk.tokenize import word_tokenize

      nltk.download('punkt')

      tokenized_data = [word_tokenize(df) for df in df]
```

```
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\naman\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

```
[45]: df
```

```
[45]:
```

	text	label
0	bynd jpmorgan reels expectations beyond meat h...	0
1	ccl rcl nomura points bookings weakness carniv...	0
2	cx cemex cut credit suisse j p morgan weak bui...	0
3	ess btig research cuts neutral https co mcyfts...	0
4	fnko funko slides piper jaffray pt cut https c...	0
...
9538	week gainers losers stoxx europe 50 dec 2 6 ec...	2
9539	tupperware brands among consumer gainers unile...	2
9540	vtv therapeutics leads healthcare gainers myom...	2
9541	work xpo pyx amkr among hour movers	2
9542	yndx qd oesx among tech movers	2

[9543 rows x 2 columns]

```
[46]: from sklearn.model_selection import train_test_split
```

```
[47]: df['text'] = df['text'].apply(lambda x: ' '.join(x) if isinstance(x, list) else
      ↪str(x))
```

```
[48]: from sklearn.feature_extraction.text import TfidfVectorizer

      tfidf = TfidfVectorizer()
```

```

        analyzer='char_wb',
        ngram_range=(3, 5),
        max_features=5000
    )

X = tfidf.fit_transform(df['text']).toarray()
y = df['label'].values

```

```
[49]: x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
↳ random_state=490)
```

```
[50]: from imblearn.over_sampling import RandomOverSampler

ros = RandomOverSampler(random_state=42)
X_resampled, y_resampled = ros.fit_resample(x_train, y_train)

print("Before:", dict(pd.Series(y_train).value_counts()))
print("After:", dict(pd.Series(y_resampled).value_counts()))
```

```

Before: {2: np.int64(4943), 1: np.int64(1525), 0: np.int64(1166)}
After: {0: np.int64(4943), 2: np.int64(4943), 1: np.int64(4943)}

```

```
[51]: import numpy as np
```

```
[52]: from sklearn.linear_model import LogisticRegression

model1 = LogisticRegression()
model1.fit(x_train, y_train)
```

```
[52]: LogisticRegression()
```

```
[53]: y_pred = model1.predict(x_test)
```

```
[54]: from sklearn.metrics import classification_report, accuracy_score

print("Accuracy:", accuracy_score(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))
```

```
Accuracy: 0.7768465165007857
```

```
Classification Report:
```

	precision	recall	f1-score	support
0	0.66	0.37	0.47	276
1	0.77	0.53	0.63	398
2	0.79	0.95	0.86	1235

accuracy			0.78	1909
macro avg	0.74	0.62	0.65	1909
weighted avg	0.77	0.78	0.76	1909

```
[55]: from sklearn.naive_bayes import MultinomialNB
```

```
model2 = MultinomialNB()
model2.fit(x_train, y_train)
```

```
[55]: MultinomialNB()
```

```
[56]: y_pred = model2.predict(x_test)
```

```
[57]: print(" Accuracy:", accuracy_score(y_test, y_pred))
print("\n Classification Report:\n", classification_report(y_test, y_pred))
```

Accuracy: 0.7464641173389209

Classification Report:

	precision	recall	f1-score	support
0	0.57	0.36	0.44	276
1	0.66	0.51	0.57	398
2	0.79	0.91	0.84	1235

accuracy			0.75	1909
macro avg	0.67	0.59	0.62	1909
weighted avg	0.73	0.75	0.73	1909

```
[58]: from sklearn.svm import LinearSVC
```

```
model3 = LinearSVC()
model3.fit(x_train, y_train)
```

```
[58]: LinearSVC()
```

```
[59]: y_pred = model3.predict(x_test)
```

```
[60]: print(" Accuracy:", accuracy_score(y_test, y_pred))
print("\n Classification Report:\n", classification_report(y_test, y_pred))
```

Accuracy: 0.7878470403352541

Classification Report:

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.63	0.51	0.56	276
1	0.73	0.61	0.66	398
2	0.83	0.91	0.87	1235
accuracy				0.79 1909
macro avg	0.73	0.68	0.70	1909
weighted avg	0.78	0.79	0.78	1909

```
[61]: from collections import Counter
```

```
[62]: # Get predictions from each model
pred1 = model1.predict(x_test)
pred2 = model2.predict(x_test)
pred3 = model3.predict(x_test)
```

```
[63]: # Majority vote for each sample
final_preds = []
for i in range(len(pred1)):
    votes = [pred1[i], pred2[i], pred3[i]]
    majority_vote = Counter(votes).most_common(1)[0][0]
    final_preds.append(majority_vote)
```

```
[64]: # Evaluate ensemble
from sklearn.metrics import accuracy_score, classification_report
print("Ensemble Accuracy:", accuracy_score(y_test, final_preds))
print("Classification Report:\n", classification_report(y_test, final_preds))
```

Ensemble Accuracy: 0.7773703509690938

Classification Report:

	precision	recall	f1-score	support
0	0.67	0.38	0.48	276
1	0.75	0.55	0.63	398
2	0.79	0.94	0.86	1235
accuracy				0.78 1909
macro avg	0.74	0.62	0.66	1909
weighted avg	0.77	0.78	0.76	1909

```
[65]: print(df['label'].value_counts())
```

```
label
2    6178
1    1923
0    1442
Name: count, dtype: int64
```