Install libraries

In [ ]:
```
%pip istall pandas
%pip install numpy
%pip install matplotlib
%pip install seaborn
%pip install joblib
```

```
ERROR: unknown command "istall" - maybe you meant "install"
Note: you may need to restart the kernel to use updated packages.
Requirement already satisfied: numpy in /home/luco/Documents/assigment/ed/
phone_reviews/venv/lib/python3.11/site-packages (1.26.4)
Note: you may need to restart the kernel to use updated packages.
Requirement already satisfied: matplotlib in /home/luco/Documents/assigmen
t/ed/phone_reviews/venv/lib/python3.11/site-packages (3.8.3)
Requirement already satisfied: contourpy>=1.0.1 in /home/luco/Documents/as
sigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotli
b) (1.2.0)
Requirement already satisfied: cycler>=0.10 in /home/luco/Documents/assigm
ent/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotlib)
(0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /home/luco/Documents/a
ssigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotl
ib) (4.50.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /home/luco/Documents/a
ssigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotl
ib) (1.4.5)
Requirement already satisfied: numpy<2,>=1.21 in /home/luco/Documents/assi
gment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotlib)
(1.26.4)
Requirement already satisfied: packaging>=20.0 in /home/luco/Documents/ass
igment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotli
b) (24.0)
Requirement already satisfied: pillow>=8 in /home/luco/Documents/assigmen
t/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotlib) (1
0.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in /home/luco/Documents/as
sigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotli
b) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in /home/luco/Document
s/assigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matpl
otlib) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in /home/luco/Documents/assigment/
ed/phone_reviews/venv/lib/python3.11/site-packages (from python-dateutil>=
2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
Requirement already satisfied: seaborn in /home/luco/Documents/assigment/e
d/phone_reviews/venv/lib/python3.11/site-packages (0.13.2)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in /home/luco/Document
s/assigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from seabo
rn) (1.26.4)
Requirement already satisfied: pandas>=1.2 in /home/luco/Documents/assigme
nt/ed/phone_reviews/venv/lib/python3.11/site-packages (from seaborn) (2.2.
1)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in /home/luco/Docum
ents/assigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from se
aborn) (3.8.3)
Requirement already satisfied: contourpy>=1.0.1 in /home/luco/Documents/as
sigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotli
b!=3.6.1,>=3.4->seaborn) (1.2.0)
Requirement already satisfied: cycler>=0.10 in /home/luco/Documents/assigm
ent/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotlib!=
3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /home/luco/Documents/a
ssigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotl
ib!=3.6.1,>=3.4->seaborn) (4.50.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /home/luco/Documents/a
ssigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotl
```

```
ib!=3.6.1,>=3.4->seaborn) (1.4.5)
Requirement already satisfied: packaging>=20.0 in /home/luco/Documents/ass
igment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotli
b!=3.6.1,>=3.4->seaborn) (24.0)
Requirement already satisfied: pillow>=8 in /home/luco/Documents/assigmen
t/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotlib!=3.
6.1,>=3.4->seaborn) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in /home/luco/Documents/as
sigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotli
b!=3.6.1,>=3.4->seaborn) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in /home/luco/Document
s/assigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matpl
otlib!=3.6.1,>=3.4->seaborn) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /home/luco/Documents/assigm
ent/ed/phone_reviews/venv/lib/python3.11/site-packages (from pandas>=1.2->
seaborn) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in /home/luco/Documents/assi
gment/ed/phone_reviews/venv/lib/python3.11/site-packages (from pandas>=1.2
->seaborn) (2024.1)
Requirement already satisfied: six>=1.5 in /home/luco/Documents/assigment/
ed/phone_reviews/venv/lib/python3.11/site-packages (from python-dateutil>=
2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
Requirement already satisfied: joblib in /home/luco/Documents/assigment/e
d/phone_reviews/venv/lib/python3.11/site-packages (1.3.2)
Note: you may need to restart the kernel to use updated packages.
```

In [ ]:
```
%pip install pandas scikit-learn
```

```
Requirement already satisfied: pandas in /home/luco/Documents/assigment/e
d/phone_reviews/venv/lib/python3.11/site-packages (2.2.1)
Requirement already satisfied: scikit-learn in /home/luco/Documents/assigm
ent/ed/phone_reviews/venv/lib/python3.11/site-packages (1.4.1.post1)
Requirement already satisfied: numpy<2,>=1.23.2 in /home/luco/Documents/as
sigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from pandas)
(1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in /home/luco/Docume
nts/assigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from pan
das) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /home/luco/Documents/assigm
ent/ed/phone_reviews/venv/lib/python3.11/site-packages (from pandas) (202
4.1)
Requirement already satisfied: tzdata>=2022.7 in /home/luco/Documents/assi
gment/ed/phone_reviews/venv/lib/python3.11/site-packages (from pandas) (20
24.1)
Requirement already satisfied: scipy>=1.6.0 in /home/luco/Documents/assigm
ent/ed/phone_reviews/venv/lib/python3.11/site-packages (from scikit-learn)
(1.12.0)
Requirement already satisfied: joblib>=1.2.0 in /home/luco/Documents/assig
ment/ed/phone_reviews/venv/lib/python3.11/site-packages (from scikit-lear
n) (1.3.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in /home/luco/Document
s/assigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from sciki
t-learn) (3.4.0)
Requirement already satisfied: six>=1.5 in /home/luco/Documents/assigment/
ed/phone_reviews/venv/lib/python3.11/site-packages (from python-dateutil>=
2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [ ]:
```
%pip install nltk
```

```
Requirement already satisfied: nltk in /home/luco/Documents/assigment/ed/p
hone_reviews/venv/lib/python3.11/site-packages (3.8.1)
Requirement already satisfied: click in /home/luco/Documents/assigment/ed/
phone_reviews/venv/lib/python3.11/site-packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in /home/luco/Documents/assigment/e
d/phone_reviews/venv/lib/python3.11/site-packages (from nltk) (1.3.2)
Requirement already satisfied: regex>=2021.8.3 in /home/luco/Documents/ass
igment/ed/phone_reviews/venv/lib/python3.11/site-packages (from nltk) (202
3.12.25)
Requirement already satisfied: tqdm in /home/luco/Documents/assigment/ed/p
hone_reviews/venv/lib/python3.11/site-packages (from nltk) (4.66.2)
Note: you may need to restart the kernel to use updated packages.
```

In [ ]:
```
%pip install matplotlib seaborn
```

```
Requirement already satisfied: matplotlib in /home/luco/Documents/assigmen
t/ed/phone_reviews/venv/lib/python3.11/site-packages (3.8.3)
Requirement already satisfied: seaborn in /home/luco/Documents/assigment/e
d/phone_reviews/venv/lib/python3.11/site-packages (0.13.2)
Requirement already satisfied: contourpy>=1.0.1 in /home/luco/Documents/as
sigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotli
b) (1.2.0)
Requirement already satisfied: cycler>=0.10 in /home/luco/Documents/assigm
ent/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotlib)
(0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /home/luco/Documents/a
ssigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotl
ib) (4.50.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /home/luco/Documents/a
ssigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotl
ib) (1.4.5)
Requirement already satisfied: numpy<2,>=1.21 in /home/luco/Documents/assi
gment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotlib)
(1.26.4)
Requirement already satisfied: packaging>=20.0 in /home/luco/Documents/ass
igment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotli
b) (24.0)
Requirement already satisfied: pillow>=8 in /home/luco/Documents/assigmen
t/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotlib) (1
0.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in /home/luco/Documents/as
sigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matplotli
b) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in /home/luco/Document
s/assigment/ed/phone_reviews/venv/lib/python3.11/site-packages (from matpl
otlib) (2.9.0.post0)
Requirement already satisfied: pandas>=1.2 in /home/luco/Documents/assigme
nt/ed/phone_reviews/venv/lib/python3.11/site-packages (from seaborn) (2.2.
1)
Requirement already satisfied: pytz>=2020.1 in /home/luco/Documents/assigm
ent/ed/phone_reviews/venv/lib/python3.11/site-packages (from pandas>=1.2->
seaborn) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in /home/luco/Documents/assi
gment/ed/phone_reviews/venv/lib/python3.11/site-packages (from pandas>=1.2
->seaborn) (2024.1)
Requirement already satisfied: six>=1.5 in /home/luco/Documents/assigment/
ed/phone_reviews/venv/lib/python3.11/site-packages (from python-dateutil>=
2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

Import libraries

In [ ]:
```
import pandas as pd
import nltk
import re
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.pipeline import Pipeline
from sklearn.metrics import classification_report, accuracy_score
import joblib
import matplotlib.pyplot as plt
import seaborn as sns
```

Download NLTK

In [ ]:
```
# nltk resources
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')
```

```
[nltk_data] Downloading package punkt to /home/luco/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /home/luco/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to /home/luco/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

Out[ ]: True

Load the dataset

In [ ]:
```
# dataset
data = pd.read_csv("./electronics_data.csv")
```

Data Preprocessing

In [ ]:
```
# data propocessing

stop_words = set(stopwords.words('english'))
lemmatizer = WordNetLemmatizer()

def preprocess_text(text):
    text = text.lower()
    text = re.sub(r'[^\w\s]', '', text)
    tokens = word_tokenize(text)
    tokens = [lemmatizer.lemmatize(word) for word in tokens if word not i
    return ' '.join(tokens)

data['clean_text'] = data['text_'].apply(preprocess_text)
```

Feature Engineering

In [ ]:
```
# Feature Engineering
X = data['clean_text']
y = data['label']
```

Split data into train and test sets

In [ ]:
```
# Split data into train and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
```

Define pipeline

In [ ]:
```
# Define pipeline
pipeline = Pipeline([
    ('tfidf', TfidfVectorizer()),
    ('clf', LogisticRegression(max_iter=1000))
])
```

Define hyperparameters grid for grid search

In [ ]:
```
# Define hyperparameters grid for grid search
param_grid = {
    'tfidf__max_features': [1000, 2000, 3000],
    'tfidf__ngram_range': [(1, 1), (1, 2)],
    'clf__C': [0.1, 1, 10]
}
```

Grid search for hyperparameter tuning

In [ ]:
```
# Grid search for hyperparameter tuning
grid_search = GridSearchCV(pipeline, param_grid, cv=5, verbose=2, n_jobs=
grid_search.fit(X_train, y_train)
```

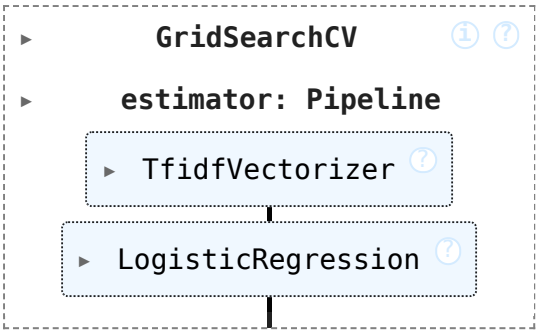
```
Fitting 5 folds for each of 18 candidates, totalling 90 fits
[CV] END clf__C=0.1, tfidf__max_features=1000, tfidf__ngram_range=(1, 1);
total time=   0.3s
[CV] END clf__C=0.1, tfidf__max_features=1000, tfidf__ngram_range=(1, 1);
total time=   0.3s
[CV] END clf__C=0.1, tfidf__max_features=1000, tfidf__ngram_range=(1, 1);
total time=   0.3s
[CV] END clf__C=0.1, tfidf__max_features=1000, tfidf__ngram_range=(1, 1);
total time=   0.3s
[CV] END clf__C=0.1, tfidf__max_features=1000, tfidf__ngram_range=(1, 1);
total time=   0.4s
[CV] END clf__C=0.1, tfidf__max_features=1000, tfidf__ngram_range=(1, 2);
total time=   0.8s
[CV] END clf__C=0.1, tfidf__max_features=1000, tfidf__ngram_range=(1, 2);
total time=   0.9s
[CV] END clf__C=0.1, tfidf__max_features=1000, tfidf__ngram_range=(1, 2);
total time=   1.0s
[CV] END clf__C=0.1, tfidf__max_features=2000, tfidf__ngram_range=(1, 1);
total time=   0.3s
[CV] END clf__C=0.1, tfidf__max_features=1000, tfidf__ngram_range=(1, 2);
total time=   1.1s
[CV] END clf__C=0.1, tfidf__max_features=2000, tfidf__ngram_range=(1, 1);
total time=   0.4s
[CV] END clf__C=0.1, tfidf__max_features=2000, tfidf__ngram_range=(1, 1);
total time=   0.3s
[CV] END clf__C=0.1, tfidf__max_features=2000, tfidf__ngram_range=(1, 1);
total time=   0.3s
[CV] END clf__C=0.1, tfidf__max_features=1000, tfidf__ngram_range=(1, 2);
total time=   0.7s
[CV] END clf__C=0.1, tfidf__max_features=2000, tfidf__ngram_range=(1, 1);
total time=   0.3s
[CV] END clf__C=0.1, tfidf__max_features=2000, tfidf__ngram_range=(1, 2);
total time=   0.8s
[CV] END clf__C=0.1, tfidf__max_features=2000, tfidf__ngram_range=(1, 2);
total time=   1.0s
[CV] END clf__C=0.1, tfidf__max_features=2000, tfidf__ngram_range=(1, 2);
total time=   0.9s
[CV] END clf__C=0.1, tfidf__max_features=2000, tfidf__ngram_range=(1, 2);
total time=   0.9s
[CV] END clf__C=0.1, tfidf__max_features=3000, tfidf__ngram_range=(1, 1);
total time=   0.3s
[CV] END clf__C=0.1, tfidf__max_features=3000, tfidf__ngram_range=(1, 1);
total time=   0.4s
[CV] END clf__C=0.1, tfidf__max_features=3000, tfidf__ngram_range=(1, 1);
total time=   0.3s
[CV] END clf__C=0.1, tfidf__max_features=3000, tfidf__ngram_range=(1, 1);
total time=   0.3s
[CV] END clf__C=0.1, tfidf__max_features=3000, tfidf__ngram_range=(1, 1);
total time=   0.3s
[CV] END clf__C=0.1, tfidf__max_features=2000, tfidf__ngram_range=(1, 2);
total time=   1.0s
[CV] END clf__C=0.1, tfidf__max_features=3000, tfidf__ngram_range=(1, 2);
total time=   0.8s
[CV] END clf__C=0.1, tfidf__max_features=3000, tfidf__ngram_range=(1, 2);
total time=   0.8s
[CV] END clf__C=0.1, tfidf__max_features=3000, tfidf__ngram_range=(1, 2);
total time=   1.1s
[CV] END clf__C=1, tfidf__max_features=1000, tfidf__ngram_range=(1, 1); to
tal time=   0.3s
[CV] END clf__C=0.1, tfidf__max_features=3000, tfidf__ngram_range=(1, 2);
```

```
total time=   1.0s
[CV] END clf__C=1, tfidf__max_features=1000, tfidf__ngram_range=(1, 1); to
tal time=   0.3s
[CV] END clf__C=1, tfidf__max_features=1000, tfidf__ngram_range=(1, 1); to
tal time=   0.3s
[CV] END clf__C=0.1, tfidf__max_features=3000, tfidf__ngram_range=(1, 2);
total time=   0.9s
[CV] END clf__C=1, tfidf__max_features=1000, tfidf__ngram_range=(1, 1); to
tal time=   0.3s
[CV] END clf__C=1, tfidf__max_features=1000, tfidf__ngram_range=(1, 1); to
tal time=   0.4s
[CV] END clf__C=1, tfidf__max_features=1000, tfidf__ngram_range=(1, 2); to
tal time=   0.9s
[CV] END clf__C=1, tfidf__max_features=1000, tfidf__ngram_range=(1, 2); to
tal time=   0.8s
[CV] END clf__C=1, tfidf__max_features=1000, tfidf__ngram_range=(1, 2); to
tal time=   1.0s
[CV] END clf__C=1, tfidf__max_features=2000, tfidf__ngram_range=(1, 1); to
tal time=   0.3s
[CV] END clf__C=1, tfidf__max_features=1000, tfidf__ngram_range=(1, 2); to
tal time=   1.1s
[CV] END clf__C=1, tfidf__max_features=2000, tfidf__ngram_range=(1, 1); to
tal time=   0.4s
[CV] END clf__C=1, tfidf__max_features=2000, tfidf__ngram_range=(1, 1); to
tal time=   0.3s
[CV] END clf__C=1, tfidf__max_features=2000, tfidf__ngram_range=(1, 1); to
tal time=   0.3s
[CV] END clf__C=1, tfidf__max_features=1000, tfidf__ngram_range=(1, 2); to
tal time=   0.9s
[CV] END clf__C=1, tfidf__max_features=2000, tfidf__ngram_range=(1, 1); to
tal time=   0.3s
[CV] END clf__C=1, tfidf__max_features=2000, tfidf__ngram_range=(1, 2); to
tal time=   0.8s
[CV] END clf__C=1, tfidf__max_features=2000, tfidf__ngram_range=(1, 2); to
tal time=   0.9s
[CV] END clf__C=1, tfidf__max_features=3000, tfidf__ngram_range=(1, 1); to
tal time=   0.3s
[CV] END clf__C=1, tfidf__max_features=2000, tfidf__ngram_range=(1, 2); to
tal time=   1.0s
[CV] END clf__C=1, tfidf__max_features=2000, tfidf__ngram_range=(1, 2); to
tal time=   1.0s
[CV] END clf__C=1, tfidf__max_features=3000, tfidf__ngram_range=(1, 1); to
tal time=   0.3s
[CV] END clf__C=1, tfidf__max_features=3000, tfidf__ngram_range=(1, 1); to
tal time=   0.3s
[CV] END clf__C=1, tfidf__max_features=2000, tfidf__ngram_range=(1, 2); to
tal time=   0.8s
[CV] END clf__C=1, tfidf__max_features=3000, tfidf__ngram_range=(1, 1); to
tal time=   0.4s
[CV] END clf__C=1, tfidf__max_features=3000, tfidf__ngram_range=(1, 1); to
tal time=   0.4s
[CV] END clf__C=1, tfidf__max_features=3000, tfidf__ngram_range=(1, 2); to
tal time=   0.8s
[CV] END clf__C=1, tfidf__max_features=3000, tfidf__ngram_range=(1, 2); to
tal time=   0.9s
[CV] END clf__C=1, tfidf__max_features=3000, tfidf__ngram_range=(1, 2); to
tal time=   0.8s
[CV] END clf__C=1, tfidf__max_features=3000, tfidf__ngram_range=(1, 2); to
tal time=   1.1s
[CV] END clf__C=10, tfidf__max_features=1000, tfidf__ngram_range=(1, 1); t
```

```
otal time=   0.3s
[CV] END clf__C=10, tfidf__max_features=1000, tfidf__ngram_range=(1, 1); t
otal time=   0.2s
[CV] END clf__C=10, tfidf__max_features=1000, tfidf__ngram_range=(1, 1); t
otal time=   0.4s
[CV] END clf__C=1, tfidf__max_features=3000, tfidf__ngram_range=(1, 2); to
tal time=   0.8s
[CV] END clf__C=10, tfidf__max_features=1000, tfidf__ngram_range=(1, 1); t
otal time=   0.3s
[CV] END clf__C=10, tfidf__max_features=1000, tfidf__ngram_range=(1, 1); t
otal time=   0.3s
[CV] END clf__C=10, tfidf__max_features=1000, tfidf__ngram_range=(1, 2); t
otal time=   0.8s
[CV] END clf__C=10, tfidf__max_features=1000, tfidf__ngram_range=(1, 2); t
otal time=   1.0s
[CV] END clf__C=10, tfidf__max_features=1000, tfidf__ngram_range=(1, 2); t
otal time=   0.7s
[CV] END clf__C=10, tfidf__max_features=1000, tfidf__ngram_range=(1, 2); t
otal time=   1.0s
[CV] END clf__C=10, tfidf__max_features=2000, tfidf__ngram_range=(1, 1); t
otal time=   0.3s
[CV] END clf__C=10, tfidf__max_features=2000, tfidf__ngram_range=(1, 1); t
otal time=   0.3s
[CV] END clf__C=10, tfidf__max_features=2000, tfidf__ngram_range=(1, 1); t
otal time=   0.4s
[CV] END clf__C=10, tfidf__max_features=2000, tfidf__ngram_range=(1, 1); t
otal time=   0.3s
[CV] END clf__C=10, tfidf__max_features=1000, tfidf__ngram_range=(1, 2); t
otal time=   0.8s
[CV] END clf__C=10, tfidf__max_features=2000, tfidf__ngram_range=(1, 1); t
otal time=   0.4s
[CV] END clf__C=10, tfidf__max_features=2000, tfidf__ngram_range=(1, 2); t
otal time=   0.8s
[CV] END clf__C=10, tfidf__max_features=2000, tfidf__ngram_range=(1, 2); t
otal time=   0.9s
[CV] END clf__C=10, tfidf__max_features=2000, tfidf__ngram_range=(1, 2); t
otal time=   1.0s
[CV] END clf__C=10, tfidf__max_features=2000, tfidf__ngram_range=(1, 2); t
otal time=   0.9s
[CV] END clf__C=10, tfidf__max_features=3000, tfidf__ngram_range=(1, 1); t
otal time=   0.3s
[CV] END clf__C=10, tfidf__max_features=3000, tfidf__ngram_range=(1, 1); t
otal time=   0.3s
[CV] END clf__C=10, tfidf__max_features=3000, tfidf__ngram_range=(1, 1); t
otal time=   0.4s
[CV] END clf__C=10, tfidf__max_features=2000, tfidf__ngram_range=(1, 2); t
otal time=   0.8s
[CV] END clf__C=10, tfidf__max_features=3000, tfidf__ngram_range=(1, 1); t
otal time=   0.4s
[CV] END clf__C=10, tfidf__max_features=3000, tfidf__ngram_range=(1, 1); t
otal time=   0.4s
[CV] END clf__C=10, tfidf__max_features=3000, tfidf__ngram_range=(1, 2); t
otal time=   0.8s
[CV] END clf__C=10, tfidf__max_features=3000, tfidf__ngram_range=(1, 2); t
otal time=   0.8s
[CV] END clf__C=10, tfidf__max_features=3000, tfidf__ngram_range=(1, 2); t
otal time=   0.9s
[CV] END clf__C=10, tfidf__max_features=3000, tfidf__ngram_range=(1, 2); t
otal time=   0.9s
```

```
[CV] END clf__C=10, tfidf__max_features=3000, tfidf__ngram_range=(1, 2); t
otal time=   0.5s
```

Out[ ]:



Best hyperparameters

In [ ]:
```
# Best hyperparameters
best_params = grid_search.best_params_
print("Best Hyperparameters:", best_params)
```

```
Best Hyperparameters: {'clf__C': 10, 'tfidf__max_features': 3000, 'tfidf__
ngram_range': (1, 2)}
```

Evaluate model

In [ ]:
```
# evaluate model
y_pred = grid_search.predict(X_test)
print("Classification Report:")
print(classification_report(y_test, y_pred))
print("Accuracy:", accuracy_score(y_test, y_pred))
```

```
Classification Report:
              precision    recall  f1-score   support

          CG       0.90      0.88      0.89       419
          OR       0.87      0.89      0.88       379

    accuracy                           0.89       798
   macro avg       0.89      0.89      0.89       798
weighted avg       0.89      0.89      0.89       798

Accuracy: 0.8859649122807017
```

Save the model

In [ ]:
```
# Save the model
joblib.dump(grid_search, 'fake_review_detection_model.pkl')
print("Model saved successfully!")
```

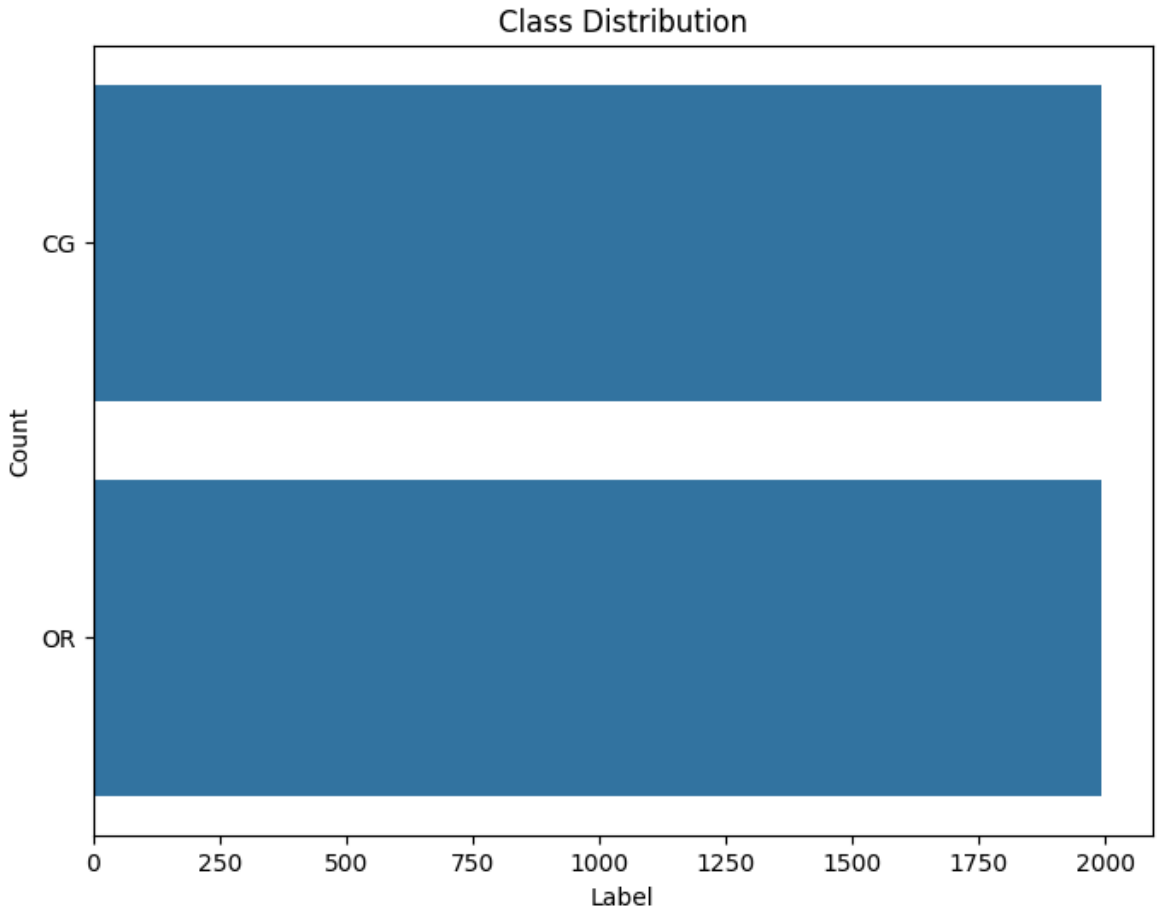
```
Model saved successfully!
```

Visualization

Visualize class distribution in the dataset

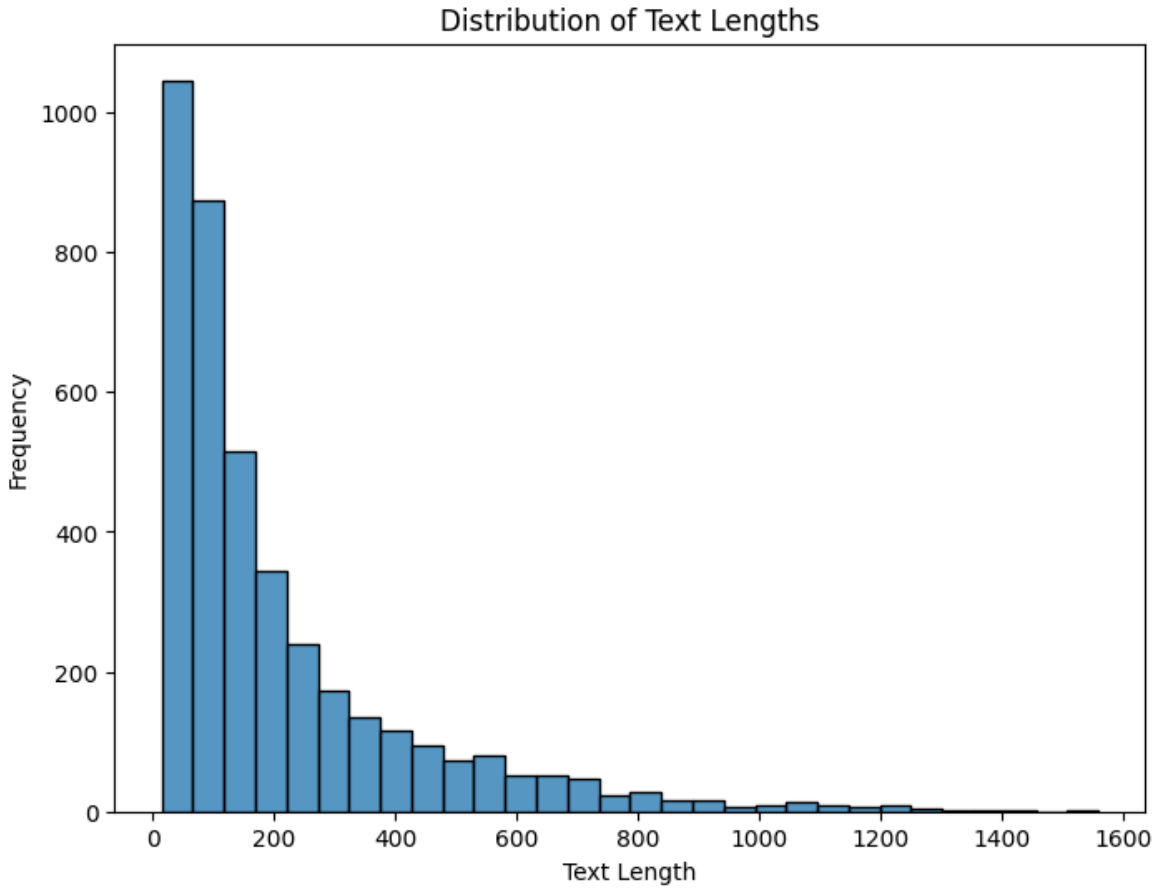
In [ ]:
```
plt.figure(figsize=(8, 6))
sns.countplot(data['label'])
plt.title('Class Distribution')
plt.xlabel('Label')
```

```
plt.ylabel('Count')
plt.show()
```



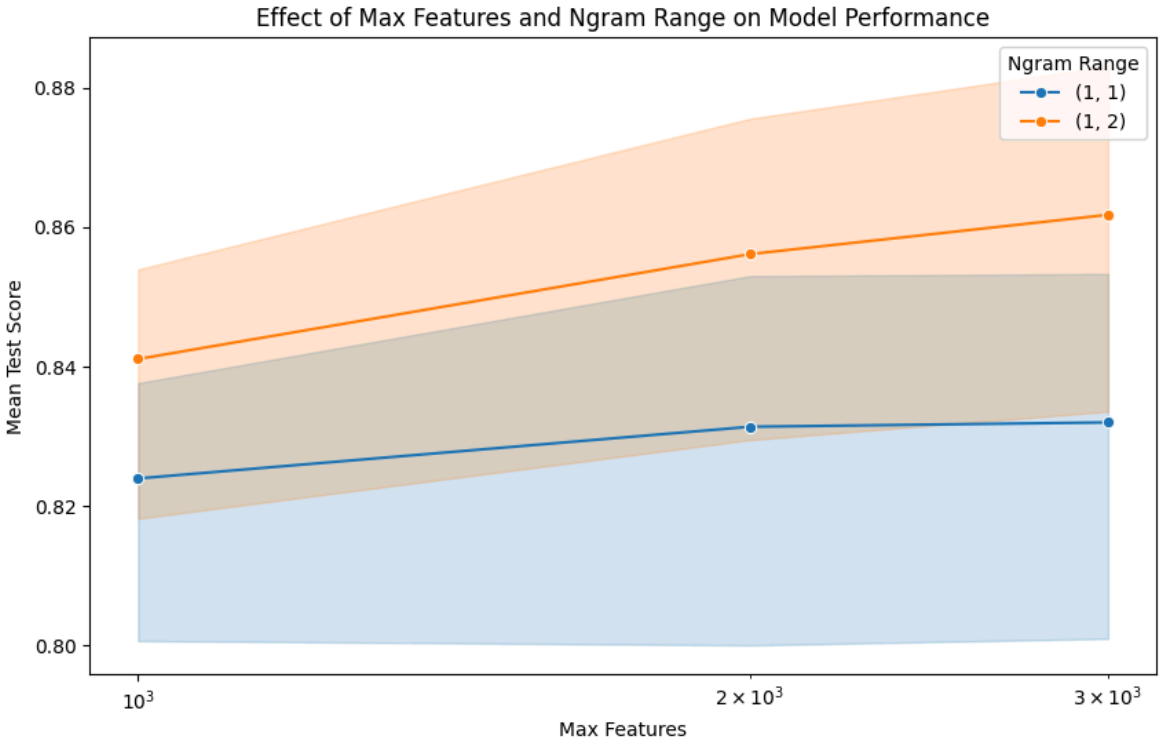
Visualize the distribution of text lengths

In [ ]:
```
text_lengths = data['clean_text'].apply(len)

plt.figure(figsize=(8, 6))
sns.histplot(text_lengths, bins=30)
plt.title('Distribution of Text Lengths')
plt.xlabel('Text Length')
plt.ylabel('Frequency')
plt.show()
```



visualize the hyperparameters' effect on model performance

In [ ]:
```
param_results = pd.DataFrame(grid_search.cv_results_)

plt.figure(figsize=(10, 6))
sns.lineplot(data=param_results, x='param_tfidf__max_features', y='mean_t
plt.xscale('log')
plt.title('Effect of Max Features and Ngram Range on Model Performance')
plt.xlabel('Max Features')
plt.ylabel('Mean Test Score')
plt.legend(title='Ngram Range')
plt.show()
```


Effect of Max Features and Ngram Range on Model Performance
Ngram Range
(1, 1)
(1, 2)
Mean Test Score
0.88
0.86
0.84
0.82
0.80
10^3
2×10^3
3×10^3
Max Features