Rakshit Awadhiya
23BCE1396
VIT, Chennai

Sentiment Analysis on Movie Reviews

```
pip install pandas scikit-learn
```

Code output: -

```
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (2.2.2)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.11/dist-packages (1.6.1)
Requirement already satisfied: numpy>=1.23.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (2.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: scipy>=1.6.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (1.15.3)
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (1.5.1)
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (3.6.0)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
```

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report
from sklearn.preprocessing import LabelEncoder

df = pd.read_csv('imdb.csv')

vec = TfidfVectorizer(stop_words='english', max_features=5000)
X = vec.fit_transform(df['review'])

le = LabelEncoder()
y = le.fit_transform(df['sentiment'])

X_tr, X_te, y_tr, y_te = train_test_split(X, y, test_size=0.2, random_state=42)

clf = LogisticRegression(max_iter=1000)
clf.fit(X_tr, y_tr)

y_pr = clf.predict(X_te)
print(classification_report(y_te, y_pr, target_names=['Negative', 'Positive']))
```

Rakshit Awadhiya
23BCE1396
VIT, Chennai

Final output: -

	precision	recall	f1-score	support
Negative	0.90	0.88	0.89	5650
Positive	0.88	0.90	0.89	5625
accuracy			0.89	11275
macro avg	0.89	0.89	0.89	11275
weighted avg	0.89	0.89	0.89	11275