Homework 2: Sentiment Classifier

The second programming assignment will familiarize you with the use of machine learning for text classification, and the use of prebuilt lexicons and bag of words to build custom features.

The primary objective for the assignment is the same as the first assignment: to predict the sentiment of a movie review. We will be providing you with the dataset containing the text of the movie reviews from IMDB, and for each review, you have to predict whether the review is positive or negative.

Data

The data is the same as the first homework. The filecontains a similar structure for the first homework.

Kaggle

As with HW1, you can make at most three submissions each day, so we encourage you to test your submission files early and observe the performance of your system. By the end of the submission period, you will have to select the two submissions, one for default and one for custom (more on this later).

Source Code

Some initial code contains methods for loading the data and lexicons, and calling the methods to run and evaluate your classifier. It also contains the code to output the submission file from your classifier (called `rf_custom_text.csv`) that you will submit to Kaggle. Your directory structure should look like this.

```
hw2
|
└──code
|    └──hw2_sentiment_classifier.ipynb
|    └──lexicon_reader.py
└──data
|    └──lexicon
|        |   inqtabs.txt
|        └──SentiWordNet_3.0.0_20130122.txt
|    └──test
|        |   0.txt
|        |   1.txt
|        |   ...
|        └──24999.txt
|    └──train
|        |   0.txt
|        |   1.txt
|        |   ...
|        └──24999.txt
|    └──train.csv
└──output
```

What to submit?

Prepare and submit a single write-up (**PDF, maximum 3 pages**) with Python source code (custom_features.py, error_analysis.ipynb, and ml_sentiment.py) compressed in a zip file to Canvas. **Do not include your student ID number** , since we might share it with the class if it's worth highlighting. The write-up pdf and code zip file should be submitted separately on Canvas. The pdf should include:

Part 1. Preliminaries, 10 points

Kaggle Team name and Kaggle accuracy of best default and custom models. The team with the best score in the competition has 10 points, the 2nd team has 9 points, the third has 8, and the others earn 7 points. You can make ***at most three*** submissions each day, so we encourage you to test your submission files early, and observe the performance of your system.

- Start with a single line header: `Id, Category`
- For each of the unlabeled speech (sorted by name) there is a line containing an increasing integer index (i.e. line number 1), then a comma, and then the string label prediction of that speech.
- See `sample_sol.csv` for example.

Part 2. Default Features, 20 points

Tune classifiers on default features (include the selected range of parameters in the code), and submit your predictions to Kaggle. Include the accuracies of the best classifier for each (LR and RF) and the two plots in the write-up, and a few sentences on how you picked the range.

Part 3. Custom Features, 35 points

Implement your custom features and vectorizers in the code, and train and tune the classifiers. Submit the predictions to Kaggle, identify the best classifier for each, and include the accuracy obtained in the report. Try at least five different sets of features and vectorizers with at least five other parameters. Include the description and comparison of your features and the vectorizers in a few sentences in the write-up.

Part 4. Analysis, 30 points

Select the best classifier with default features and the best classifier with custom features. The analysis will focus on comparing these two classifiers. First, use eli5 to generate the global importance weights for both classifiers in the notebook (show_weights), and in a few sentences in the write-up, describe what is different between them. Then, in the notebook, generate 2 examples each where the classifiers disagree: (i) positive reviews, where only the default is correct, (ii) positive reviews, where only the custom is correct, (iii) negative reviews, where only the default is correct, and (iv) negative reviews, where only the custom is correct. Also, include 2 examples each of when both classifiers are correct, and both are incorrect. In the write-up, include a paragraph describing the insights these errors provide about the differences between the classifiers, especially the advantages/disadvantages of your custom features.

Part 5. Statement of Collaborations, 5 points

It is mandatory to include a Statement of Collaboration in each submission with respect to the guidelines below. Include the names of everyone involved in the discussions (especially in-person ones) and what was discussed. All students are required to follow the academic honesty guidelines posted on the course website. For programming assignments, in particular, we encourage the students to organize (perhaps using Piazza) to discuss the task descriptions, requirements, bugs in our code, and the relevant technical content before they start working on it. However, you should not discuss the specific solutions, and, as a guiding principle, you are not allowed to take anything written or drawn away from these discussions (i.e., no photographs of the blackboard, written notes, referring to Piazza, etc.). Especially after you have started working on the assignment, try to restrict the discussion to Piazza as much as possible, so that there is no doubt as to the extent of your collaboration.

In [1]:

```
import sys
import os
import csv
import pickle
import eli5
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from sklearn.base import BaseEstimator, TransformerMixin
from sklearn.feature_extraction import DictVectorizer
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
from sklearn.pipeline import FeatureUnion, make_pipeline

import lexicon_reader
```

In [17]:

```
class Dataset:
    def __init__(self, data, start_idx, end_idx):
        self.data = data
        self.reviews = [row['Review'] for row in data[start_idx:end_idx]]
        self.labels = [row['Category'] for row in data[start_idx:end_idx]]
        self.vecs = None

def get_training_and_dev_data(filedir, dev_rate=0.2):
    with open(os.path.join(filedir, 'train.csv'), 'r', encoding='utf-8') as c
        data = [row for row in csv.DictReader(csvfile, delimiter=',')]
        for entry in data:
            with open(os.path.join(filedir, 'train', entry['FileIndex'] + '.t
                entry['Review'] = reviewfile.read()
    dev_idx = int(len(data) * (1 - dev_rate))
    return Dataset(data, 0, dev_idx), Dataset(data, dev_idx, len(data))

def get_test_data(filedir, output_file_name):
    testfiledir = os.path.join(filedir, 'test')
    with open(output_file_name, 'w', encoding='utf-8') as csvfile:
        writer = csv.DictWriter(csvfile, delimiter=',', fieldnames=['Id', 'Ca
        writer.writeheader()
        for filename in sorted(os.listdir(testfiledir), key=lambda x: int(os.
            with open(os.path.join(testfiledir, filename), 'r', encoding='utf
                fileindex = os.path.splitext(filename)[0]
                review = reviewfile.read()
                yield (fileindex, review)

def write_predictions(filedir, classifier, output_file_name):
    with open(output_file_name, 'w', encoding='utf-8') as csvfile:
        writer = csv.DictWriter(csvfile, delimiter=',', fieldnames=['Id', 'Ca
        writer.writeheader()
        for (fileindex, review) in get_test_data(filedir, output_file_name):
            prediction = dict()
            prediction['Id'] = fileindex
            prediction['Category'] = classifier.predict([review])[0]
            writer.writerow(prediction)

def get_trained_classifier(data, model, features):
    ppl = make_pipeline(features, model)
    return ppl.fit(data.reviews, data.labels)

def get_custom_features(filedir):
    return FeatureUnion([
        ('custom_feats', make_pipeline(CustomFeats(filedir), DictVectorizer()
        ('bag_of_words', get_custom_vectorizer())
    ])

def save(classifier, filedir, output_file_path):
    with open(output_file_path + ".pkl", 'wb') as f:
        pickle.dump(classifier, f)
    write_predictions(filedir, classifier, output_file_path + "_test.csv")

def load_classifier(input_file_path):
    return pickle.load(open(input_file_path, 'rb'))

def plot(xs, train_accuracy_list, dev_accuracy_list, output_file_path=None):
```

```
    plt.clf()
    plt.plot(xs, train_accuracy_list, label='train')
    plt.plot(xs, dev_accuracy_list, label='dev')
    plt.ylabel('Accuracy')
    plt.legend()
    if output_file_path is not None:
        plt.savefig(output_file_path)
    else:
        plt.show()
```

In [18]:

```
# Load data

filedir = '../data'
print("Reading data")
train_data, dev_data = get_training_and_dev_data(filedir)
```

```
Reading data
```

Part 2. Default Features

We are providing code for training a machine learning classifier for sentiment classification using unigrams as features, i.e. `CountVectorizer()`. The first goal is to optimize the hyper-parameters of logistic regression by modifying `get_tuned_lr`. The regularization weight C is the primary hyper-parameter for logistic regression. Currently, the range for the parameter is `np.arange(0.5, 3.5, 0.5)` but this should be modified. When running this function, you will see both training and dev accuracy printed. There will also be a plot `lr.png` that will be saved that you can view. Based on these, adjust the range for C.

The next goal is to optimize the parameters for random forest. To do so, you need to modify `get_tuned_rf`. `n_estimators` is the parameter of interest used by random forest. Currently, the range is set to `np.arange(5, 35, 5)` but this should also be modified. Like before, when running this function, you will see both training accuracy and dev accuracy, and the plot `rf.png` will be saved. Based on what you see, adjust the parameters accordingly.

Running `save(tuned_lr, filedir, 'lr_default')` will save the classifier as `lr_default.pkl` which you will need for your error analysis. It will also run the classifier on the test set and save the results as `lr_default_test.csv`, which you can upload to Kaggle. Similarly, the next line will output the files `rf_default.pkl` and `rf_default_test.csv`.

In [4]:

```
def get_tuned_lr(train, dev, features, output_file_path='./lr.png'):
    train_vecs = features.fit_transform(train.reviews)
    dev_vecs = features.transform(dev.reviews)
    train_accuracy_list = list()
    dev_accuracy_list = list()

    # ----------------------------------------------------------------------
    # TODO: You will change this range, or may want to use np.logspace instea
    cs = np.arange(0.1, 3.5, 0.1)

    for c in cs:
        model = LogisticRegression(C=c)
        model.fit(train_vecs, train.labels)
        train_preds = model.predict(train_vecs)
        dev_preds = model.predict(dev_vecs)
        (train_score, dev_score) = (accuracy_score(train.labels, train_preds)
        print("Train Accuracy:", train_score, ", Dev Accuracy:", dev_score)
        train_accuracy_list.append(train_score)
        dev_accuracy_list.append(dev_score)
    plot(cs, train_accuracy_list, dev_accuracy_list, output_file_path)
    best_model = LogisticRegression(C=cs[np.argmax(dev_accuracy_list)])
    return get_trained_classifier(train, best_model, features)
```

In [5]:

```
def get_tuned_rf(train, dev, features, output_file_path='./rf.png'):
    train_vecs = features.fit_transform(train.reviews)
    dev_vecs = features.transform(dev.reviews)
    train_accuracy_list = list()
    dev_accuracy_list = list()

    # ----------------------------------------------------------------------
    # TODO: You will change this range, and try different parameters to tune
    n_estimators = np.arange(20,100,20)

    for num_estimator in n_estimators:
        model = RandomForestClassifier(n_estimators=num_estimator)
        model.fit(train_vecs, train.labels)
        train_preds = model.predict(train_vecs)
        dev_preds = model.predict(dev_vecs)
        (train_score, dev_score) = (accuracy_score(train.labels, train_preds)
        print("Train Accuracy:", train_score, ", Dev Accuracy:", dev_score)
        train_accuracy_list.append(train_score)
        dev_accuracy_list.append(dev_score)
    plot(n_estimators, train_accuracy_list, dev_accuracy_list, output_file_pa
    best_model = RandomForestClassifier(n_estimators=n_estimators[np.argmax(d
    return get_trained_classifier(train, best_model, features)
```

In [6]:

```
# Some example code to get a trained classifier
print("Training model")
lr_with_default = get_trained_classifier(train_data, LogisticRegression(), Cc

# You can see some of the predictions of the classifier by running the follow
print(lr_with_default.predict(["This movie sucks!", "This movie is great!"]))

# You can then experiment with tuning the classifiers
# Experiment with the parameters in the get_tuned_lr and get_tuned_rf methods
# Look at the files lr.png and rf.png that are saved after running each of th
print("Tuning model")
tuned_lr = get_tuned_lr(train_data, dev_data, CountVectorizer())

# After playing with the parameters and finding a good classifier, you can sa
# This will save the classifier to a pickle object which you can then load la
# As well as this will run the classifier on the test set which you can then
print("Saving model and predictions")
save(tuned_lr, filedir, 'lr_default')
```

```
Training model

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


['0' '1']
Tuning model

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.9704 , Dev Accuracy: 0.8914

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

```
Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.9817 , Dev Accuracy: 0.891

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.98595 , Dev Accuracy: 0.889

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.9797 , Dev Accuracy: 0.889

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.98755 , Dev Accuracy: 0.887
```

```
C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.98655 , Dev Accuracy: 0.887

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.9861 , Dev Accuracy: 0.8858

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.98775 , Dev Accuracy: 0.8862

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logist
ic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown i
n:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://s
cikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-re
gression (https://scikit-learn.org/stable/modules/linear_model.html#logis
```

```
tic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.98805 , Dev Accuracy: 0.8852

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.98905 , Dev Accuracy: 0.8858

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.9915 , Dev Accuracy: 0.8876

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.986 , Dev Accuracy: 0.8856

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
```

```
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.99125 , Dev Accuracy: 0.885

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.99195 , Dev Accuracy: 0.8842

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.99205 , Dev Accuracy: 0.8858

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.99045 , Dev Accuracy: 0.8834

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

```
Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.99095 , Dev Accuracy: 0.8834

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.99125 , Dev Accuracy: 0.8824

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.99195 , Dev Accuracy: 0.884

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.9913 , Dev Accuracy: 0.884
```

```
C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.99145 , Dev Accuracy: 0.8788

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.99055 , Dev Accuracy: 0.8832

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.99085 , Dev Accuracy: 0.881

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logist
ic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown i
n:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://s
cikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-re
gression (https://scikit-learn.org/stable/modules/linear_model.html#logis
```

```
tic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
```

```
Train Accuracy: 0.98945 , Dev Accuracy: 0.8838
```

```
C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
```

```
Train Accuracy: 0.9911 , Dev Accuracy: 0.883
```

```
C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
```

```
Train Accuracy: 0.9916 , Dev Accuracy: 0.8838
```

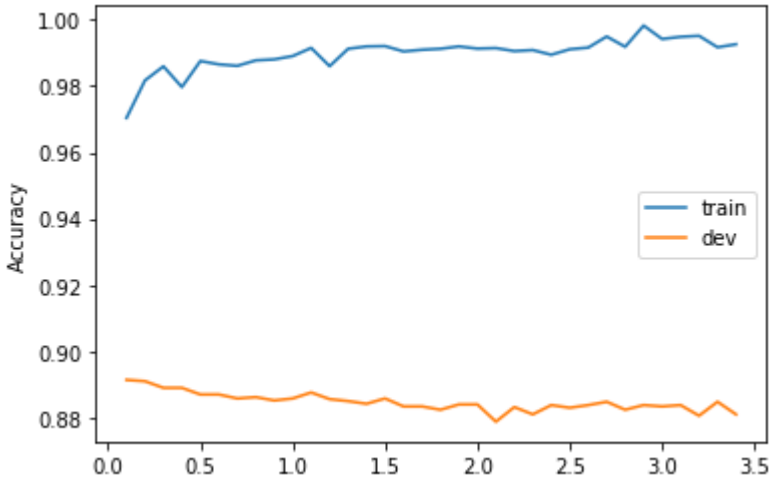
```
C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
```

```
Train Accuracy: 0.99495 , Dev Accuracy: 0.8848
```

```
C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
```

```
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.99185 , Dev Accuracy: 0.8824

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.99825 , Dev Accuracy: 0.8838

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.9942 , Dev Accuracy: 0.8834

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.99485 , Dev Accuracy: 0.8838

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

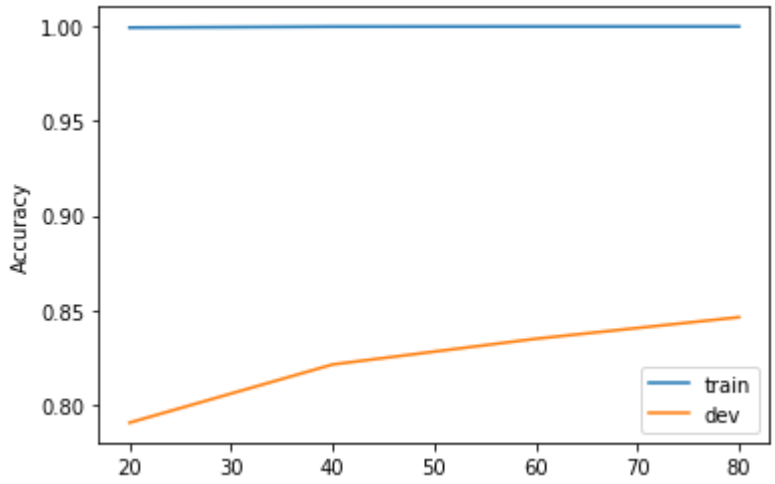
```
Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.99515 , Dev Accuracy: 0.8806

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.9917 , Dev Accuracy: 0.8848

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Train Accuracy: 0.9926 , Dev Accuracy: 0.881

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,


Saving model and predictions
```


1.00
0.98
0.96
0.94
0.92
0.90
0.88
Accuracy
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
train
dev

In [7]:

```
# Some example code to get a trained classifier
print("Training model")
rf_with_default = get_trained_classifier(train_data, RandomForestClassifier()

# You can see some of the predictions of the classifier by running the follow
print(rf_with_default.predict(["This movie sucks!", "This movie is great!"]))

# You can then experiment with tuning the classifiers
# Experiment with the parameters in the get_tuned_lr and get_tuned_rf methods
# Look at the files lr.png and rf.png that are saved after running each of th
print("Tuning model")
tuned_rf = get_tuned_rf(train_data, dev_data, CountVectorizer())

# After playing with the parameters and finding a good classifier, you can sa
# This will save the classifier to a pickle object which you can then load la
# As well as this will run the classifier on the test set which you can then
print("Saving model and predictions")
save(tuned_rf, filedir, 'rf_default')
```

```
Training model
['0' '1']
Tuning model
Train Accuracy: 0.99945 , Dev Accuracy: 0.7904
Train Accuracy: 0.99995 , Dev Accuracy: 0.8212
Train Accuracy: 1.0 , Dev Accuracy: 0.8348
Train Accuracy: 1.0 , Dev Accuracy: 0.8462
Saving model and predictions
```



Part 3. Custom Features and Vectorizers

The next goal for the assignment is to improve upon these classifiers by introducing your own features. Similar to the in-class activity in week 4, you will design features that utilize lexicons and regular expressions, etc., and `experiment with the vectorizer` (e.g., unigrams vs. bigrams, counts vs. TF-IDF, etc.). To implement these `features`, you will need to modify the features function, and you can change the `vectorizer`.

As before, tune the parameters for both logistic regression and random forest, but this time with your custom features.

Run the save methods to save the classifiers (`lr_custom.pkl` and `rf_custom.pkl`) and predictions (`lr_custom_text.csv` and `rf_custom_text.csv`), and upload the latter files to Kaggle.

In [10]:

```
class CustomFeats(BaseEstimator, TransformerMixin):
    """Extract features from each document for DictVectorizer"""
    def __init__(self, filedir):
        self.feat_names = set()
        lexicon_dir = os.path.join(filedir, 'lexicon')
        self.inqtabs_dict = lexicon_reader.read_inqtabs(os.path.join(lexicon_
        self.swn_dict = lexicon_reader.read_senti_word_net(os.path.join(lexic

    def fit(self, x, y=None):
        return self

    @staticmethod
    def word_count(review):
        words = review.split(' ')
        return len(words)

    def pos_count(self, review):
        words = review.split(' ')
        count = 0
        for word in words:
            if word in self.inqtabs_dict.keys() and self.inqtabs_dict[word] =
                count += 1
        return count

    def neg_count(self, review):
        words = review.split(' ')
        count = 0
        for word in words:
            if word in self.inqtabs_dict.keys() and self.inqtabs_dict[word] =
                count += 1
        return count

    def features(self, review):
        return {
            # ------------------------------------------------------------------
            # 4 example features
            # TODO: Add your own here e.g. word_count, and pos_count
            'length': len(review),
            'num_sentences': review.count('.'),
            'num_words': self.word_count(review),
            'pos_count': self.pos_count(review),
            'neg_count': self.neg_count(review)
        }

    def get_feature_names(self):
        return list(self.feat_names)

    def transform(self, reviews):
        feats = []
        for review in reviews:
            f = self.features(review)
            [self.feat_names.add(k) for k in f]
            feats.append(f)
        return feats
```

```
def get_custom_vectorizer():
    # ----------------------------------------------------------------------
    #TODO: Experiment with different vectorizers
    return CountVectorizer()
    #return TfidfVectorizer()
    #return CountVectorizer(ngram_range = (2,2))
    #return TfidfVectorizer(ngram_range = (3,3))
```

In [11]:

```
# Experiment with different features by modifiying custom_features.py and tes
# (Again, you can look at the lr.png  and rf.png that are saved after running
print("Tuning model")
tuned_lr = get_tuned_lr(train_data, dev_data, get_custom_features(filedir))
tuned_rf = get_tuned_rf(train_data, dev_data, get_custom_features(filedir))

print("Saving model and predictions")
save(tuned_lr, filedir, 'lr_custom')
save(tuned_rf, filedir, 'rf_custom')
```

```
Tuning model

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8056 , Dev Accuracy: 0.8042

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.80375 , Dev Accuracy: 0.8022

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.83485 , Dev Accuracy: 0.8248

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
```

```
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8339 , Dev Accuracy: 0.824

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8125 , Dev Accuracy: 0.8082

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8263 , Dev Accuracy: 0.8158

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.81075 , Dev Accuracy: 0.8058
```

```
C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
Train Accuracy: 0.83615 , Dev Accuracy: 0.828

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.82435 , Dev Accuracy: 0.8142

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.82655 , Dev Accuracy: 0.8136

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.83985 , Dev Accuracy: 0.831
```

```
C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8356 , Dev Accuracy: 0.8262

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.82945 , Dev Accuracy: 0.819

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.83255 , Dev Accuracy: 0.825

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logist
ic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown i
n:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://s
cikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-re
gression (https://scikit-learn.org/stable/modules/linear_model.html#logis
tic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
```

```
Train Accuracy: 0.8275 , Dev Accuracy: 0.8146

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8318 , Dev Accuracy: 0.819

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8195 , Dev Accuracy: 0.8098

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.82645 , Dev Accuracy: 0.8172

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logist
ic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown i
n:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://s
cikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-re
gression (https://scikit-learn.org/stable/modules/linear_model.html#logis
```

```
tic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8286 , Dev Accuracy: 0.8166

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.815 , Dev Accuracy: 0.8046

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8329 , Dev Accuracy: 0.8218

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8118 , Dev Accuracy: 0.8018

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logist
ic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown i
n:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://s
cikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
```

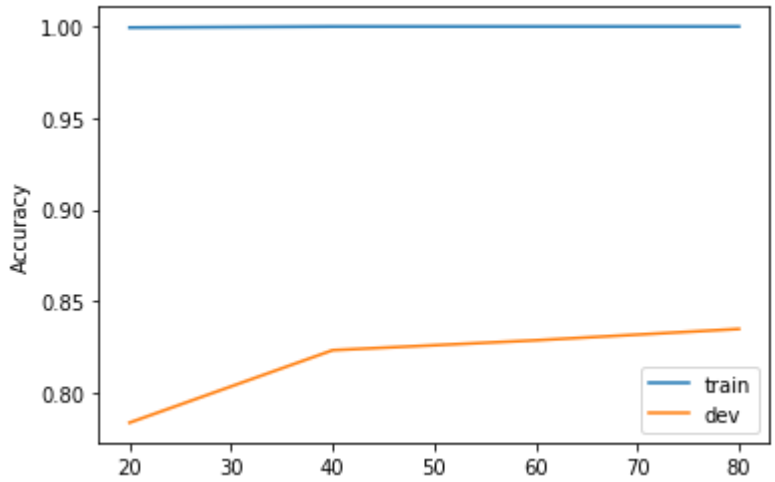
```
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8242 , Dev Accuracy: 0.8142

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8021 , Dev Accuracy: 0.8014

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.81825 , Dev Accuracy: 0.8062

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8299 , Dev Accuracy: 0.8188

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://s
```

```
cikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-re
gression (https://scikit-learn.org/stable/modules/linear_model.html#logis
tic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.82385 , Dev Accuracy: 0.8152

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8217 , Dev Accuracy: 0.8136

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.83345 , Dev Accuracy: 0.8214

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logisti
c.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sci
kit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regr
ession (https://scikit-learn.org/stable/modules/linear_model.html#logistic-
regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.80725 , Dev Accuracy: 0.8026

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logist
ic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown i
```

```
n:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
Train Accuracy: 0.8342 , Dev Accuracy: 0.826

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.80435 , Dev Accuracy: 0.8014

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.81755 , Dev Accuracy: 0.8104

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,

Train Accuracy: 0.8254 , Dev Accuracy: 0.814

C:\Users\student\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:818: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

```
Increase the number of iterations (max_iter) or scale the data as shown i
n:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://s
cikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-re
gression (https://scikit-learn.org/stable/modules/linear_model.html#logis
tic-regression)
  extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG,
```

```
Train Accuracy: 0.99945 , Dev Accuracy: 0.7836
Train Accuracy: 1.0 , Dev Accuracy: 0.8232
Train Accuracy: 1.0 , Dev Accuracy: 0.8286
Train Accuracy: 1.0 , Dev Accuracy: 0.8348
Saving model and predictions
```



Part 4. Error Analysis

Along with tuning classifiers and designing features to achieve as high of accuracy as possible, you also need to perform an analysis of the classifier in this assignment. For this analysis, modify the function to get metrics and feature weights for all four of your best classifiers (note, it is currently set up to only compute metrics for two, so you will need to either modify it to get metrics for all four or run it twice). You may need to update `load_classifier('lr_default.pkl')` replacing `lr_default.pkl` with the model you have (the one that was saved when you run the saved method earlier).

The rest of the notebook also contains the two primary approaches for analysis that use the `eli5` package: (1) global importance weights of individual classifiers, and (2) weights of individual words, for example, predictions. The notebook also includes code for easily comparing classifiers to each other.

In [22]:

```
pd.set_option('display.max_colwidth', -1)

def get_error_type(pred, label):
    # return the type of error: tp,fp,tn,fn
    if pred == label:
        return "tp" if pred == '1' else "tn"
    return "fp" if pred == '1' else "fn"

# Change this for your different classifiers
classifier1 = load_classifier('lr_default.pkl')
classifier2 = load_classifier('rf_default.pkl')
classifier3 = load_classifier('lr_custom.pkl')
classifier4 = load_classifier('rf_custom.pkl')

# Create pandas dataframe
predictions = pd.DataFrame.from_dict(dev_data.data)

# Classify data points using classifier1
predictions['Classifier1Prediction'] = classifier1.predict(predictions['Revie
predictions['Classifier1ErrorType'] = predictions.apply(lambda row: get_error

# Classify data points using classifier 2
predictions['Classifier2Prediction'] = classifier2.predict(predictions['Revie
predictions['Classifier2ErrorType'] = predictions.apply(lambda row: get_error

# Classify data points using classifier 3
predictions['Classifier3Prediction'] = classifier3.predict(predictions['Revie
predictions['Classifier3ErrorType'] = predictions.apply(lambda row: get_error

# Classify data points using classifier 4
predictions['Classifier4Prediction'] = classifier4.predict(predictions['Revie
predictions['Classifier4ErrorType'] = predictions.apply(lambda row: get_error

# Get metrics for each classifier
def print_metrics(error_type_counts):
    accuracy = (error_type_counts['tp'] + error_type_counts['tn']) / sum(erro
    precision = error_type_counts['tp'] / (error_type_counts['tp'] + error_ty
    recall = error_type_counts['tp'] / (error_type_counts['tp'] + error_type_
    print("Accuracy:", accuracy, "\nPrecision:", precision, "\nRecall:", reca

print("Classifier1 Metrics")
print_metrics(predictions['Classifier1ErrorType'].value_counts())
print("\nClassifier2 Metrics")
print_metrics(predictions['Classifier2ErrorType'].value_counts())
print("\nClassifier3 Metrics")
print_metrics(predictions['Classifier3ErrorType'].value_counts())
print("\nClassifier4 Metrics")
print_metrics(predictions['Classifier4ErrorType'].value_counts())
```

```
C:\Users\student\anaconda3\lib\site-packages\ipykernel_launcher.py:1: Futur
eWarning: Passing a negative integer is deprecated in version 1.0 and will
not be supported in future version. Instead, use None to not limit the colu
mn width.
  """Entry point for launching an IPython kernel.
```

```
Classifier1 Metrics
Accuracy: 0.9546
Precision: 0.9534753810549836
Recall: 0.95584
F1: 0.9546562262794135

Classifier2 Metrics
Accuracy: 0.96924
Precision: 0.9674822666772934
Recall: 0.97112
F1: 0.9692977202858625

Classifier3 Metrics
Accuracy: 0.83808
Precision: 0.8276476973174135
Recall: 0.854
F1: 0.8406173714465706

Classifier4 Metrics
Accuracy: 0.96872
Precision: 0.9683453237410072
Recall: 0.96912
F1: 0.968732506997201
```

In [23]:
```
eli5.show_weights(classifier1, top=25)
```

Out[23]:

y=1 top features

Weight[?]	Feature
+0.782	excellent
+0.725	perfect
+0.715	funniest
+0.701	superb
+0.682	refreshing
… 34420 more positive …	
… 33966 more negative …	
-0.668	lame
-0.675	lacks
-0.684	unfortunately
-0.716	worse
-0.722	poor
-0.729	laughable
-0.730	dull
-0.733	disappointing
-0.735	ridiculous
-0.740	save
-0.742	badly
-0.757	avoid
-0.795	mess
-0.844	horrible
-0.881	boring
-0.954	poorly
-1.031	disappointment
-1.087	awful
-1.198	worst
-1.336	waste

In [25]:
```
eli5.show_weights(classifier2, top=25)
```

Out[25]:

Weight	Feature
0.0105 ± 0.0297	bad
0.0093 ± 0.0233	worst
0.0071 ± 0.0169	great
0.0049 ± 0.0140	waste
0.0047 ± 0.0141	awful
0.0042 ± 0.0048	and
0.0034 ± 0.0078	no
0.0031 ± 0.0024	the
0.0030 ± 0.0073	nothing
0.0030 ± 0.0083	boring
0.0029 ± 0.0082	terrible
0.0029 ± 0.0075	acting
0.0028 ± 0.0080	excellent
0.0027 ± 0.0025	of
0.0027 ± 0.0022	is
0.0026 ± 0.0069	money
0.0026 ± 0.0057	just
0.0025 ± 0.0073	minutes
0.0025 ± 0.0058	wonderful
0.0025 ± 0.0054	plot
0.0025 ± 0.0044	best
0.0025 ± 0.0026	this
0.0025 ± 0.0027	was
0.0024 ± 0.0077	worse
0.0024 ± 0.0021	in
… 68385 more …	

In [26]:

```
# See some examples of errors for each classifier
# Modify the code to get false negatives and errors for Classifier2)
# Classifier 1
predictions[predictions['Classifier1ErrorType'] == 'fp'].sample(10)

# See where they disagree
# Modify the code to find cases where one classifier's prediction is correct
predictions['ClassifiersAgree'] = predictions['Classifier1Prediction'] == pre
disagreements = predictions[predictions['ClassifiersAgree'] == False]
print("# Cases where the two classifiers disagree:", len(disagreements), "->"
disagreements.sample(10)
```

```
# Cases where the two classifiers disagree: 1252 -> 5.008 %
```

Out[26]:

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Cla

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Cla
22550	22550	0	A never ending frenzy of clever visual ironies does not necessarily create an engaging film. The "Blonde Wig" half of the movie never took off perhaps due to too much self-indulgence by its makers. The Wong Faye half (featuring a very playful, if Karen Carpenter looking, Faye Wong) holds much more appeal. All the ingredients are there, however, the girl-meets-boy story element takes a back seat to artsy cleverness. Character development is uneven. Emotion is missing. For music lovers, Wong Faye's "Mung Jung Yun" (Cantonese version of the Cranberry's smash hit "Dreams") is used effectively in Chungking Express. Faye Wong also recorded a Mandarin language version called "Zhen Tuo." Both are on CD, although only "Mung Jung Yun" is found on the official movie soundtrack CD.	0	tn	

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Cla
502	502	1	Surprisingly not terrible and well animated for one of Disney's straight to video throw away sequels. Like the previous sequel (The Lion King 2) I was glad that Disney brought back most of the original voice actors which makes a big difference and they kept a good level of traditional animation. The plot wanders around for a while but we are distracted by an unending string of jokes ranging from hilarious to dull. To break up the detached plot and jokes they gave us some silly musical sequences, which much like the jokes, range from entertaining to a quick trip to the fridge. For the most part the MST3K-like moments are bland and full of untapped potential and really don't add a whole lot to the movie other than to act as a vehicle for an hour-long flashback. The new characters are at least likable, and the old characters are out doing their thing so I can't fault them there. Overall this movie in not bad and it makes for a nice frivolous filler between the more serious Lion King titles.	0	fn	

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Cla
6649	6649	1	My main problem with the film is that it goes on too long. Other then that, it's pretty good. Paul Muni plays a poor Chinese farmer who is about to get married through an arranged marriage. Luise Rainer is a servant girl who gets married to Muni. They live with Muni's father on a farm and they are doing pretty bad. When he finally gets some money to buy some more land, a drought hits and nothing is growing. Everybody stars to head north by Muni stays behind at first. When they leave and arrive at town they find that their are no jobs and they are worse off than before. They even think about selling their youngest daughter as a slave for some money but decide against it. When a bunch of people start looting the town, the military show up and start executing people . Paul Muni does a good job and Luise Rainer won a second oscar for this movie.	0	fn	

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Cla
22714	22714	0	I really have problems rating this movie. It is directed brilliantly, there is obviously a lot of money in it. Gere and Danes are intense (although her screen personality could use a bit more defining and spicing up), editing and cinematography are excellent. On the other hand, it is one of those really really sick movies where one cannot help but wonder whether the director himself likes to stage specific scenes, and, yes, one cannot help but wonder how many copycats will such a movie inspire. In purely artistic terms, it is a 9, but I really have to ask myself who these people are giving their money to produce such a movie	0	tn	

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Cla
11937	11937	0	This movie proves that you can't judge a movie by the awesome artwork on the DVD cover. It also goes to show that you should learn more about a movie before you buy it (or get it for someone at Christmas). The beginning of this movie actually looks somewhat promising. Well, until you meet the characters. Pumpkin Jack (the old guy from down the street) brings the college co-eds a book full of witch's spells that he leaves at their annual haunted house (where the movie takes place). After that there is some drinking, fighting, and soft core porn. Then the action of the movie finally takes place after over an hour. Overall, Hallow's End was predictable, unsuspensful, and reminiscent of a soft-core porn. This movie is probably best viewed with a group of friends who have nothing better to do, as it is a good movie to make fun of. And for first-time viewers, it is really fun making predictions of the order of people who die.	1	fp	

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Cla
6911	6911	0	This movie is AWESOME. I watched it the other day with my cousin Jay-Jay. He said it was alright, but i think it RULEZZZ! I mean, it's so cool. Ted V. Mikels is so brave and smart. He made a movie totally unlike those terrible Hollywood films, like the Matrix and STop or my Mom will Shoot. It could have been better, though. I like ninjas and pirates. I also like that big talon that the funny man wears. I think he's the coolest guy since that Domino Pizza claymation guy. Not only does this movie look really cool, like those out-of-focus movies my dad made of my birthday when I turned 6. BUt it tells a complex tale with dozens of characters that seem to be totally unrelated, but they all meet up in the end. It's genius how this web is woven to make everything meet up. I wish Ted V. Mikels would make a sequel. But it needs more aliens. And a pirate.	1	fp	

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Cla
20778	20778	0	I generally LIKE watching Burt Lancaster's films--especially when he is needed to go nuts with his imposing screen presence like in Elmer Gantry. However, his greatest strength, his magnetism, was occasionally also his greatest weakness as he rarely, if ever, underplayed ANYTHING. And it is this lack of subtlety that really hinders The Rainmaker. Now I understand that his character was meant to be a sort of showman but how Katherine Hepburn could fall under his spell is completely inexplicable. She is supposed to be smart but doesn't seem so when Lancaster's blarney is being thrown about the screen! In addition to this, the story is perhaps one of the most stagy looking films I have ever seen and it is way too obvious that this is a movie based on a play. It just looks like it was mostly filmed in a sound stage instead of in the great wide open West like it was supposed to be. Overall, a very overrated film.	0	tn	

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Cla
12122	12122	0	I thought it was not the best re-cap episode I've every seen (though my viewing partner handed me a tissue in anticipation of the Brendan Fraser moment...*sigh*). It was nice to see Cox outside of the incessantly brittle "Coxism State" he is in these days, if only for brief moments. I also enjoyed trying to place the episodes included by the length of the character's hair (or height, in case of JD) and the youthfulness of the earliest episodes. I can also see how Zach might be well on the way to a very Chevy Chase/or is that Matthew Perry? prat-fall induced chemical slide (already acknowledged on Conan). A little side note, the song (now stuck in my head) from the janitor-induced dance montage was "Diner" by Martin Sexton.	1	fp	
21823	21823	0	This beautifully filmed and scripted episode was let down for two reasons. 1) Perhaps it was the morality of the 1950s talking, but no man left alone on an asteroid for years would react with such hysterical negativity to the gift of a female	0	tn	

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Cla
		android. 2) It wasn't an android at all, but a woman, the beautiful Jean Marsh. The popularity of the sex doll industry in the coming decades could have traced its origins back to this episode if they'd done it properly. In fact, the modernization of sex-bots are in the news as I speak. Robots were not new to movies or television when this episode was made, so they could have at least had her act like one. Her fleshiness would then have added a creepy element. Instead, it becomes a nice little love story about two humans on faraway star. The Twilight Zone always stretched the imagination and credulity. Normally no one cared. But this episode seemed hamstrung by a Calvinist morality eschewing what would have amounted to masturbation with a machine, or downright carelessness.			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Cla
6766	6766	0	Even though an animated film it really bored everyone under at least 6. As a grown up who grew up in an area with wild horses and native americans, it felt this was a combination of PC mixed in with too many fantasy films created by people who never lived in the area they filmed about. Talk to those who have lived on horse back, most treat their animals like family members, regardless of background. Regardless of background we have dealt with good and bad breakers of wild horses. I had to explain that was a real life issues to us vs the movie makers views to children who were surprised to see how PC showed a world different than what they knew in reality. This dreamworks break from the normal disney or dreamworks fare of cute talking animals burning up the screen was nice from the older viewer point of view. But if you live in an area similar to what is shown, you may end up answering questions.	1	fp	

In [27]:

```
# See some examples of errors for each classifier
# Classifier2
predictions[predictions['Classifier2ErrorType'] == 'fp'].sample(10)

# See where they disagree
# Modify the code to find cases where one classifier's prediction is correct
predictions['ClassifiersAgree'] = predictions['Classifier2Prediction'] == pre
disagreements = predictions[predictions['ClassifiersAgree'] == False]
print("# Cases where the two classifiers disagree:", len(disagreements), "->"
disagreements.sample(10)
```

```
# Cases where the two classifiers disagree: 3893 -> 15.572 %
```

Out[27]:

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
4031	4031	0	I was one of the few non-liberals who showed up to see Steve's video. It was quite an experience... in propaganda film-making and boredom. I was hoping the film might be an actual documentary of Michael Moore's visit to my local school, UVSC, but it turned out to be another liberal, slash-and-burn effort to slam conservatives and the local religious community. It sure seems self-serving for a filmmaker to make a documentary that only reflects his preconceptions on issues. What's more surprising is to see all the '10' votes his homeys have posted here. Did they even see the video? Golly gee Batman, this must rank with All The President's Men! Their ratings are as obvious as the bias in this film. Yeah, like stacking the votes at IMDb will help a lame movie. Maybe my vote will help balance this out.	0	tn	

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
15145	15145	1	Moonchild is a very difficult movie to categorise. It's easiest to think of it as several snapshots of the lives of the two central characters. The fact that these characters are members of a street gang set in an multicultural city of the near future and that one of them is a vampire does not preclude them from having moments like any other people, and this is one of the places where this movie is different to anything else I've ever heard of. It doesn't get wrapped up in the fact that one of the main characters is a vampire, it's just something that has to be dealt with like any other problem. The way the characters interact is surprisingly realistic- there are embarrassing relatives and tricks that are meant to look cool that just don't work, which leaves the film with a lovely sense of not taking itself too seriously for the most part. The other area that really stood out to	1	tp	

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
		me is the languages. The fictional city of Mallepa contains various cultural groups, and characters speak the language that they would be expected to speak. Japanese gang members speak Japanese to each other, but Chinese when talking to characters of Chinese descent. Possibly the most amusing exchange involves an Australian and is conducted in English. The actors of the four arguably main characters have three separate mother tongues between them and speak varying levels of each others' languages, so it's quite a feat that the movie was made at all. Which, I suppose, brings me to the lead actors. Much has been made of the fact that the movie stars two of Japan's biggest rockstars, Gackt and Hyde, as well as Taiwanese superstar Lee-hom Wang, whether it is to praise them for their acting or criticise it or			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
			simply fangirl about them. In my opinion, Lee-hom is the best at playing a straight and realistic character. However, any lack of acting ability on Gackt's part is mostly masked by the fact that the character he plays is prone to being over-dramatic. I wasn't sure if Hyde's character was supposed to be as sulky and sarcastic as he came across, but it doesn't really detract from the movie either way. There are several scenes which take rather melodramatic turns, which made it difficult for them to affect me much emotionally (Although this doesn't seem to stop a lot of people). I found it's best to just enjoy the movie for what it is and not take it too seriously- It's perfect for getting out and watching with a group of friends. It does have its flaws, but overall it was very enjoyable and I'd highly recommend it to anyone who doesn't mind a few subtitles.			
22227	22227	1	It's been so long since I've seen this	1	tp	

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
		movie (at least 15 years) and yet it still haunts me with a vivid image of the horrific consequences that prisoners of war can face despite the terms of the Geneva Convention. A unit of Australian underwater demolitions experts are captured in an archipelago near Japan following a successful mission to set mines in a Japanese harbor. Once in prison these men expect the same treatment as any other POWs but to their dismay soon learn from a friendly Japanese prison guard that they are being tried as spies since they were out of uniform when captured. The consequences of such an infraction, by Japanese martial code, is execution by beheading. Despite their pleas, and the pleas of the sympathetic prison guard, the day of reckoning approaches like a ticking time bomb. The tension is so high you will actually hear the ticking,			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
			though it may just be your chest pounding with the percussion of a marching execution squad. The ending is actually too painful to reenact in my head much less write it here. But I can promise you--you'll never forget it. Good luck finding the video in the U.S.			
8527	8527	1	Superbly trashy and wondrously unpretentious 80's exploitation, hooray! The pre-credits opening sequences somewhat give the false impression that we're dealing with a serious and harrowing drama, but you need not fear because barely ten minutes later we're up until our necks in nonsensical chainsaw battles, rough fist-fights, lurid dialogs and gratuitous nudity! Bo and Ingrid are two orphaned siblings with an unusually close and even slightly perverted relationship. Can you imagine playfully ripping off the towel that covers your sister's naked body and then	1	tp	

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
		stare at her unshaven genitals for several whole minutes? Well, Bo does that to his sister and, judging by her dubbed laughter, she doesn't mind at all. Sick, dude! Anyway, as kids they fled from Russia with their parents, but nasty soldiers brutally slaughtered mommy and daddy. A friendly smuggler took custody over them, however, and even raised and trained Bo and Ingrid into expert smugglers. When the actual plot lifts off, 20 years later, they're facing their ultimate quest as the mythical and incredibly valuable White Fire diamond is coincidentally found in a mine. Very few things in life ever made as little sense as the plot and narrative structure of "White Fire", but it sure is a lot of fun to watch. Most of the time you have no clue who's beating up who or for what cause (and I bet the actors understood even less) but whatever! The violence is magnificently			

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
		grotesque and every single plot twist is pleasingly retarded. The script goes totally bonkers beyond repair when suddenly and I won't reveal for what reason Bo needs a replacement for Ingrid and Fred Williamson enters the scene with a big cigar in his mouth and his sleazy black fingers all over the local prostitutes. Bo's principal opponent is an Italian chick with big breasts but a hideous accent, the preposterous but catchy theme song plays at least a dozen times throughout the film, there's the obligatory "we're-falling-in-love" montage and loads of other attractions! My God, what a brilliant experience. The original French title translates itself as "Life to Survive", which is uniquely appropriate because it makes just as much sense as the rest of the movie: None!			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
21484	21484	1	This film is wonderful in every way that modern action adventures are not. Take some time. Relax, enjoy. Think. People who see this movie as slow or plodding or dull really need to take a week off and watch it several times until their short attention span mind comes to grips with the possibility of being involved with a cause or even beautiful story in a beautiful place for no other reason than because it isn't hurrying to make the points you so emphatically need it to make in the short time alloted. At first I was apprehensive of Brosnan playing a native American. Given the story line though, I think it was apt casting. Now, back to my hermiting. - Jahfre	1	tp	
18394	18394	0	Maddy (Debbie Rochon) is a mentally unstable young woman with a troubled past who gets more than she bargained for when she goes to a pool party with a handsome coworker. When her date and his friends	0	tn	

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
		jokingly say they belong to a `Murder Club,' Maddy takes it seriously and moves straight up to `Level 3' by bashing in the brains of a woman in a parking garage (for denting her car!). But is Maddy also the one donning a plastic mask and killing off other members of the group or has someone else lost it? The plot of this film (originally titled MAKE 'EM BLEED) is very poorly conceived, full of holes and spirals completely out of control before a ludicrous, out-of-left-field twist ending. Some of the dialogue is downright laughable. I didn't have a problem with Rochon's performance, but the supporting cast was atrocious. However, I managed to sit through this Full Moon release thoroughly entertained. There's plenty of skin and blood and it's the perfect type of flick to sit around with a group of your buddies and pick apart. Horror fans may also enjoy the cameos from Brinke			

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
		Stevens and Lloyd Kaufman (as Debbie's parents) and Julie Strain (an early victim). Score: 4 out of 10			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
20360	20360	0	The film starts to slowly when we got to the cinema we thought it looked quite good but after about 5 mins we were all bored out of our minds and wondering what kind of film we had come to see, i don't like this film and wouldn't recommend it to anyone, the best part of the night was when the alarm and lights came back on because the project broke down because we thought we could all go home. this has to be one of the worst films i have ever seen we were all bored out of or minds and most of the people in the cinema actually RAN out of the doors at the end because it was so rubbish. i am surprised that no one walked out earlier than that. if you go and see it make sure you something to keep you busy, better still Don't go and see it at all.	1	fp	
14966	14966	1	A series of random, seemingly insignificant thefts at her sister's boarding house has	1	tp	

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
		Miss Lemon quite agitated. A ring, light bulbs, a rucksack, a lighter, a stethoscope, a shoe there seems to be no rhyme or reason to any of it. Miss Lemon asks her employer, the great Belgian detective Hercule Poirot, to look into the matter. But what Poirot sees is something far more sinister than Miss Lemon could have imagined. And Poirot's fears are confirmed when one of the students living in the boarding house if found murdered. It's up to Poirot to bring a killer to justice. Hickory Dickory Dock is a solid, but not spectacular, entry in the long running Poirot series. I appreciate how faithful the script is to Agatha Christie's original story. I realize that certain liberties had to be taken, but I appreciate the effort nonetheless. The major points of the mystery are all there the petty thefts, the boarding house, the students, the			

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
		ripped rucksack, and, of course, Poirot's ability to see something sinister going on before it actually happens. With a few exceptions, the cast of students is almost as I pictured them. Damian Lewis and Jessica Lloyd standout among the group. As mush as I always enjoy David Suchet's Poirot, I get a real kick out of the episodes with Phillip Jackson's Inspector Japp and Pauline Moran's Miss Lemon. This episode is a real treat as Miss Lemon gets more screen time than usual. Finally, I enjoyed the use of the ever present mouse as an observer of the activities in the hostel. It's a fun little play on the Hickory Dickory Dock title. I realized while re-watching Hickory Dickory Dock just what a tremendous influence Agatha Christie's work was on the highly stylized Italian mystery films, or Gialli, of the 60s and 70s. Take the murder of Mrs.			

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
		Nicoletis as an example. If you were to bump up the graphic nature of the scene, you would have something straight out of an early 70s Giallo. In fact, the entire plot of Hickory Dickory Dock could have been used in a Giallo. It's just convoluted and interesting enough to have worked.			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
18286	18286	0	This movie wasn't awful but it wasn't very good. I am a big fan Toni Collette I think she is a very beautiful and talented actress. The movie starts off about Robin Williams who is a writer and gets a book from a 14 year old kid. The book is great and he cant't believe a kid wrote it. Toni Collette plays the kids guardian who you don't know if this kid really exists or if she's making it all up. I am not gonna ruin the movie but I will say this the movie is not scary. The acting is pretty good and Toni Collette's performance was awesome as well as Robin Williams. The movie was a huge disappointment in my opinion I would wait for it to come to DVD.	0	tn	
16423	16423	0	This grainy film has a cult following and one of those word-of-mouth features you just had to see. Maybe hard to believe, but there is a rural community in southwest Arkansas, Fouke, that knows the	0	tn	

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifier
		legend is true. This tale is told in documentary-style narrated by Vern Stierman and filmed in actual locations talking to actual folks involved. The legend changes with the telling, but during the late 60s and most of the 70s the surrounding area of Fouke was visited by a Bigfoot-like creature that traveled along Boggy Creek. Long limbed with three toes and standing over 7 foot tall, this hirsute creature periodically caused damage and frightened the 'bejeebers' out of most of the community. I personally crossed over the small Boggy Creek bridge in 1974, and yes the hair on the back of my neck did rise. Of course it was about 1 a.m. in the rain. By the time I arrived in Shreveport, I was laughing.			

In [30]:

```
# See some examples of errors for each classifier
# Classifier3
predictions[predictions['Classifier3ErrorType'] == 'fp'].sample(10)

# See where they disagree
# Modify the code to find cases where one classifier's prediction is correct
predictions['ClassifiersAgree'] = predictions['Classifier3Prediction'] == pre
disagreements = predictions[predictions['ClassifiersAgree'] == False]
print("# Cases where the two classifiers disagree:", len(disagreements), "->"
disagreements.sample(10)
```

```
# Cases where the two classifiers disagree: 3936 -> 15.744 %
```

Out[30]:

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
4660	4660	0	This documentary begins with an interesting premise -- it makes an intriguing and convincing argument that the history of Jesus as is commonly believed is probably a myth. Sadly, though, after priming us with this, the movie completely shifts gears and becomes little more than a non-stop attack on Christianity, and pretty much focusing on the easy targets. The writer/director clearly has some issues with the Church (he is a former evangelical Christian and has some legit anger) and this film seems to be his form of release. It'd be interesting to see the first 20 minutes expanded, but as a whole, the movie is disappointing.	0	tn	
14058	14058	0	The main reason people still care about "Carlton-Browne Of The F.O." is that it	0	tn	

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
		features Peter Sellers in a second-billed role. But watching this film to see Peter Sellers is a mistake. Sellers plays Amphibulos, a vaguely reptilian prime minister of the dirt-poor island nation of Gaillardia, formerly a British colony, now hosting a lot of Russian diggers during the height of the Cold War. Amphibulos wants to play both U.K. and Soviet interests against each other for easy profit, "everything very friendly and all our cards under the table". Terry-Thomas is the title character, a lazy British diplomat anxious to show Gaillardia that Great Britain hasn't forgotten them, all appearances to the contrary. A positive review here says: "The reason this movie is considered average is because the comedy is understated." I would argue that the reason "Carlton-Browne" is considered below average is because the comedy is non-existent. After a decent opening that establishes the film's only two strengths, a sympathetically doltish Terry-Thomas and John Addison's full-on larky score, things quickly slow down			

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
		into a series of slow burns and lame miscommunication jokes. The low opinion of Carlton-Browne by his boss and the obscurity of Gaillardia (which no one can find on a map) is milked to death. By the time we actually reach the island (after a labored series of airsick jokes), expectations are quite low. They're still too high, though. The island itself, which seems to exist either in Latin America or the Mediterranean, is so pathetic its honor guard faints at the airport, and the review stand falls apart in the middle of a parade. The army is apparently still horse drawn, allowing for another lame aural gag by a thick-accented announcer: "In war, the army uses many horse." Sellers never quite takes center stage even when we're on his character's island. The plot is taken over instead by Ian Bannen as King Loris, who inherits the throne of Gaillardia after his father's assassination. Bannen is dull and plays his part as straight as it is written. Normally this would make him the likely target for scene-stealing by Sellers, but trapped behind a thick accent and greasy			

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
		moustache, Sellers is only a threat to those of us who remember him far more happily in two other films made this same year, "The Mouse That Roared" and "I'm All Right, Jack." Strange that this film, like "Jack", was a Boulting Brothers production, with Roy Boulting here serving as co-director alongside Jeffrey Dell. Usually Boulting films combine wicked social satire with anything-goes comedy, but here there are only fey jabs in either direction. Amphibulos works his mangled-English vibe for all its worth ("This man is like, how do you say, the bull in the Chinese ship") while Carlton-Browne is generally ragged on by his superior far more than he seems to deserve. The weakest and most protracted element of the film is young Loris's romance with Ilyena. Score one point for her being played by ravishing Luciana Paluzzi, dock one for the fact that they are apparently cousins is never addressed. The film winds up with a lamely staged revolution whose surprise resolution will surprise no one, and a final bit of action by Carlton-Browne			

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
		that would seem to nail the lid on his coffin literally. Apparently he lives to see another day, but the film of the same name is strictly DOA.			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
558	558	0	It's impossible for me to objectively consider this movie. Not that I haven't tried, mind you - but I sit down, and I pop in the aged VHS, and I watch the opening...and suddenly I'm five years old again and clutching my very own Care Bear and watching the movie with open eyes and an eager heart. I can see, objectively, that this movie is a BIZARRE combination of cuddly baby merchandising-mascots and creepy prepubescent children with evil powers that has a thin story and uninteresting animation. But my inner five-year-old goes, "Yay! Care Bears!" every time I think about it. So - I'd only (cautiously, reluctantly) recommend this movie for those who saw it during their early youth and can call on the awesome power of nostalgia while watching it (like me) OR those lovably cynical Gen-X/Y-ers who deliberately seek out the wonderfully bad/strange (a category in which this movie...definitely belongs). To those actually looking for a compelling movie or wholesome family entertainment: You might want to keep looking.	0	tn	

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
22178	22178	1	This film illustrates the worst part of surviving war, the memories. For many soldiers, men and women alike, returning home can be the beginning of real problems. I am reminded of my father and his brothers returning from WWII. For one of my uncles the war was never over. He survived the D-Day invasion, something akin to the first 20 minutes of Saving Private Ryan. For him the memories not only lingered but tortured him. He became an alcoholic as did several of my cousins, his sons. Jump ahead 60 years and place the soldiers in a different war, in a different country, the result is the same. When I saw this at the KC FilmFest, I was reminded that there are somethings about war that never change. The idealistic young men and women are not spared the emotional torment of what happened in Iraq, and especially if you are against the war you will come away with more compassion for the soldiers there trying to do what they believe or have been told is right. The tag line from the Vietnam war film Platoon says it all. "The First Casualty of War is Innocence."	1	tp	

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
10938	10938	1	"Pet Sematary" is an adaptation from the Stephen King novel of the same title. The story follows the Creeds - an all American, middle-class family, who move into a house out in the country. The family consists of Louis and Rachel, and their two young children, Ellie and their toddler son, Gage. The house couldn't be better, and the family meets a strange but friendly old man, Jud, who lives across the road. He leads them down an old path into the woods one day where a pet graveyard lies - filled with a huge amount of animal graves. And just beyond there, lies a sacred Indian burial ground that seems to possess a strange power. When the family cat, Church, is killed, Louis sees it fit to bury him in the pet cemetery - and strangely enough, soon after, Church returns to life. But there's something evil about him now, he isn't the same cat he used to be. And when a tragic accident takes the life of young Gage, Louis decides to apply the same concept in hopes of reviving his dead son... unfortunately, he gets more than he bargained for. Having read Stephen King's novel, I can say that the book is much better	1	tp	

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
		than the film. Not to say the movie is bad, because it isn't - the book is just a little bit better. The real strength in this film lies in it's story, which is both bizarre but extremely original, something that King's stories are typically known for. The script is very well adapted from the story, and while it minorly differs in some aspects, it's a pretty good page-to-screen transformation. There are a few plot holes here and there, nothing major though. Besides that, this movie is actually pretty scary, and it succeeds in it's intention to do so. There are some really disturbing scenes throughout the film, and I'd have to say that the flashback sequence of Rachel's sister Zelda is the number one. Honestly, that is one of the most disgusting, disturbing things I've ever seen in a horror film - it's not gory and bloody, it's just flat out sickening. One thing's for sure, that image won't leave your head anytime soon. The performances in this film were all very up to par and I really had no problem there. This film is actually on the gory side, there are plenty of nasty little sequences to please all of the gore hounds out			

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
		there, including the shocker of an ending. I really liked the way they ended the film, it was abrupt and somewhat inconclusive, but it worked better that way with all things considered. Overall, "Pet Sematary" is a good horror movie that I'd recommend to those who are fans of either Stephen King or just fans of the genre in general. The story is the film's greatest asset and it's a creepy one too. One of the better Stephen King adaptations I'd say. 7/10.			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
17903	17903	0	This is an action Western. James Steart leads an all star cast in the scenic Northwest, which is filmed in great splendor. The scenery and costumes are great. There is action and adventure. Stewart plays a wealthy cattleman who runs afoul of a crooked government in the old Nothwest. The main drawback is the stereotypical cynic that Hollywood has always made into a hero. Even when this movie was made, the cynic was the stereotypical hero, and the one Stewart portrays really has few saving graces. He is kind to his two partners, and that does give him an extra dimension of credibility and likability. However, he is so piggish to everyone else, it is hard to really care for him, or to accept him. He is much like the one dimensional spaghetti Western characters (cut not that bad). Still, the minor characters are quite enjoyable. Walter Brennan, Royal Dano, Harry Morgan, and others make this worth watching.	1	fp	

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
14927	14927	1	I found this film to be the usual French slap in America's face. The camera, all too often, focuses on fat people, on sloppy homes and on tacky rural areas. While the narration seems to sympathize with and admire the small town folks who are introduced to the viewer, the cinematography exploits and demeans them. There were, undoubtedly, thin people to be seen in Glencoe and neat, organized homes, but Malle chose to show us the worst of what was there to be seen. I can only hope that some American filmmakers will go to France to reveal to the American public its worst elements. I can assure you, as a frequent visitor to France, that all is not well there. Foreign immigrants are not readily assimilated, thus creating severe social inequities. But Americans are not eager to unmask the French for their prejudice toward their own compatriots and their envy toward the U.S., so we're not likely to see films on the subject.	1		tp

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
24560	24560	1	I had no idea what this film was about or even knew that it existed until about 1 month ago when I stumbled upon when I was searching for other films that stared Dominic Monaghan. I thought this film was a strange insight into the mind of a none sleeper and what his/her mind may be going through in the hours that they spend awake when the rest of the world around them is asleep,it was an interesting film and a good part was played by Dom.....I believe that even though this film you cannot buy anywhere (well I've never seen it anywhere) you must see it if you ever get the chance because it will really make you think about those people around us that cannot sleep and have to suffer night after night of not been able to sleep or only get about 1 hour of sleep every night so overall it was an interesting film of good substance.	1	tp	
16717	16717	0	This is the first Guinea Pig film from Japan and this is the sickest, in my opinion. A bunch of guys torture a girl for several days before finally killing her. And at this point, I will say that these films are NOT real! They are	0	tn	

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
		faked horror films which try to be as realistic as possible. The scenes are sickening but also unrealistic in many cases. For example, when they kick the girl in the floor, we can clearly see how they kick and stump the floor near the girl! And how stupid this looks! The sound effects are also unrealistic and don't make sense. Other scenes include animal intestines thrown on the girl, the girl exposed to loud noises for many hours, the ripping off of fingernails, worms placed on the wounds in the girl's body, the eye pierced and mutilated in horrific detail and stuff like that. Very sick and mean spirited film and has absolutely nothing valuable or cinematically significant. This first entry is the sickest and most amateurish Guinea Pig, although it is not as bloody as the next part, Flowers of Flesh and Blood, which tries to be as shocking as possible. Guinea Pig: Devil's Experiment is perhaps the sickest thing I've seen and the closest thing to snuff there is. This is still (of course) faked s(n/t)uff, the only difference to genuine "snuff film" is that no one dies or hurts for real in this film. I cannot recommend this to			

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
		anyone since thi s is so s****y and repulsive. They who consider this is a great horror film understand nothing about cinema and the real meaning of it. I watched this as a curiosity (as the other parts in the series) and now I know how insignificant trash these are. They work only in shock level and that's not too valuable cinematic achievement. Devil's Experiment is perhaps the sickest film I've seen and Mermaid in a Manhole (Guinea Pig 4) is perhaps the most disgusting film I've seen. So these are pretty extreme in my book, but that's all they are.			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classi
12444	12444	1	I just saw this movie today with my children (son, 10 and daughter, 4.5) at the 3rd Annual Roger Ebert Overlooked Film Festival. After the film the children in the audience were allowed to ask questions to the Director, Tian-Ming Wu. He (through a translator) told several stories about his life and the making of the film. All tangents aside, both of my children really enjoyed this movie. Of course, I had to paraphrase many of the subtitles for my daughter, but much of the film is visually self-explanatory. I won't give anything away, but the bottom line is that this film is SO MUCH better than 95% of the Hollywood crap (especially children's films) out there. Cheers. p.s. There is a "real"/original King of Masks who can/could do 12 masks at once. The actor in the movie trained and learned to do up to 4 masks at a time (then they would cut and change to 4 new masks).	1	tp	

In [29]:

```
# See some examples of errors for each classifier
# Classifier4
predictions[predictions['Classifier4ErrorType'] == 'fp'].sample(10)

# See where they disagree
# Modify the code to find cases where one classifier's prediction is correct
predictions['ClassifiersAgree'] = predictions['Classifier1Prediction'] == pre
disagreements = predictions[predictions['ClassifiersAgree'] == False]
print("# Cases where the two classifiers disagree:", len(disagreements), "->"
disagreements.sample(10)
```

```
# Cases where the two classifiers disagree: 1265 -> 5.06 %
```

Out[29]:

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
23789	23789	0	The film starts with a voice over telling the audience where they are, and who the characters are. And that is the moment i started to dislike the movie. With all the endless possibilities any film director have in hand, i really find it a very easy and cheap solution to express the situation with a voice over telling everything. I actually believe voice overs are betrayals to the film making concept. I hate to hear from a voice over saying where we are, which date we are at, and especially what the characters feel and think. I believe that a director has to find a visual way to transmit the feelings and the thoughts of the characters to the audience.	0	tn	

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
		 But after the bad influencing intro, a very striking movie begins and keeps going for a fairly long enough time. The lives of a middle class family and all the members individually are depicted in a perfect realistic way. I think the director has a talent for capturing real life situations. For example, a father who has to make his private calls from the bathroom might seem abnormal at first, but life itself leads us some situations which might seem abnormal but also very normal as well. I think the director is a very good observer about real life. But that is it. After a while the realism in the movie begins to sacrifice the story-telling. I really felt like I'm having a big headache because of the non-stop talking characters. It was as if the actors and actresses were given the subject and were allowed to improvise the dialogs. It is realistic really, but characters always asking "really, is that so" etc. to each other, or			

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
		characters saying "no" or "are you listening to me," ten times when saying it only once is just enough causes me to have a headache. I also think the play practicing and book reading scenes are more then they should be. I understand that the play and the book in the movie are very much related to the plot, but i think the director has missed the point where he should stop showing these scenes.			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
23769	23769	1	When I went to watch this movie my expectations were really low, but I was pleasantly surprised. I thought I was going to watch a boring teen-flick, BUT in fact the plot is interesting and well executed, the acting was somewhat convincing - especially from Melville who really shows his talent in this movie, and the fight scenes were - for a low budget movie - very well done . I think this movie deserves a broader audience than it has received. It is a movie, which can be seen by the whole family - maybe not the smallest of kids, since it contains some rather rough scenes. A movie about love, and the problems that can occur, when you go against your family traditions. Yes, the movie is very much like "Bend it like Beckham", but I actually think this movie pulls it off better.	1	tp
367	367	1	Well, well....Roeg touched a bit of a nerve there, didn't he? He was a genius	0	fn

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
		while he was cataloguing his various characters' descents into psychosis for a couple of decades, but as soon as he has the bad taste to suggest that redemption (or even some good advice) might be found in the bad old Catholic church, the hipper-than-thou alternative movie crowd gets extra vicious. Worse still, Theresa Russell's character - faced with experiences that nothing in her avowedly rationalist outlook has an explanation for, is unwillingly forced to deal with those experiences on another level - that of the spiritual. You know, the realm of the ignorant and superstitious, the sort of thing that the art-house cinephiles are supposed to be above. Oh, the horror... So she finds her marriage - the idea that it might be a uniquely important commitment - affirmed by what seems uncomfortably like divine intervention. People who find this idea prima facie offensive could			

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
		maybe ask themselves why they instinctively jump into attack mode at being challenged to take seriously the idea of a spiritual dimension to their lives. But they probably won't. Sure, this film has some problems, notably Talia Shire's delirious hamwork as the overwrought nun, 1950s-style attire and all. And the dialogue between Marie Davenport and the young priest in their last scene is straight out of the Spellbound School of Glib Interpretations (though Hitchcock's movie escaped similar charges due to the source of wisdom having impeccably secular credentials as a Freudian psychoanalyst). But, sadly, Nicolas Roeg appears to have copped a critical mauling as much for even asking the question as for the possible answers this film presents.			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
22444	22444	1	I wasn't going to watch this show. But, I'm glad I did. The critics of this just don't get it! It's one of the funniest and most entertaining thing on T.V at the present moment! Though, when the interviews were done with common folks they probably seemed useless; but, put them in the mouth of animals and insects, and it's a laugh riot. I laughed so hard, I had tears in my eyes. The pig with the babies suckling and her mother is priceless. The husband and wife birds talking about health problems, and the male bird taking a crap after the wife said she was constipated completely broke me up! Creature Comforts is the most imaginative show I've ever seen in awhile! Hopefully, it will be back next summer when this run is over.	1	tp
22072	22072	0	I anticipated the release of the film as much as any fan of the Broadway play. I waited and read reviews for months about the award winning	0	tn

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
		performances. I mean with the star power of Eddie Murphy, Jamie Foxx, Beyonce Knowles, Danny Glover... the movie couldn't be less than 4 out of 4 stars, right? WRONG! I was definitely disappointed by the finished product. The film did not match up to the publicity hype it was given and the only saving graces were Eddie Murphy, Anika Noni Rose and Jennifer Hudson. Eddie Murphy's James Brownesque performance rescues the movie just when it hits its multiple lulls and Jennifer Hudson's performance compels you to pay attention each time she's on screen. Her performance of "And I Am Telling You" was the only time that I felt the hype was deserved. You cringed as she begged her no good man to let her stay in the group and in his life. As many reviewers have stated, she steals the movie from the more experienced actors and deserves all the accolades she's receiving for this			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
			performance. Anika Noni Rose was also a strong presence with a great voice and comedic talent. Jamie Foxx and Beyonce Knowles, on the other hand, cruised through their performances. Foxx's acting skills for this film seemed to predate his extraordinary "Ray" performance and Beyonce Knowles was on an extended fashion photo shoot or video taping, posing and shimmying her way through the movie. Her performance wasn't strong enough to make you care about her character at any point in the film. The movie was too hyped, 30 minutes and 1 song (Beyonce's "heartfelt" solo to Jamie Foxx) too long. DH -- Vancouver, WA			
5713	5713	1	From a modern sensibility, it's sometimes hard to watch older films. It's annoying to have to watch the stereotypical wallflower librarian have to take off her glasses and become pretty and stupid to win a man.	0		fn

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
		Especially such a shallow and inconstant man. He's obviously a player (I wouldn't trust him to stay true to her) who doesn't want to settle down, who only looks at dumb attractive women and always calls them "baby" (ick!). Even after she totally changes her appearance and her life for him, he only goes to her after he's (supposedly) rejected by another woman and learns that Connie spent all her money renovating a boat for him. I wanted her to stand up to him, not pathetically chase after him! His sudden conversion within a few minutes was totally unrealistic and did not work for me. Apart from that subplot, I did like the movie. How can you not like sailors dancing with each other?! (You can tell they were from San Francisco.... ;D) The "rehearsal" dance was great, watching Ginger Rogers purposefully fall in and out of the "correct steps" was great. The last			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
			dance scene "Face the Music" with the beautiful costumes and the art deco set was beautiful. And I really enjoyed "We Saw the Sea" (though they did use it a few too many times, as if they realized it was their best song). Anyway, the plot was a bit weak, like most musicals (IMO) - and the songs were OK, but the dancing was worth watching the film for. I wish they could have showed some shots of San Francisco since that was were the film was supposedly set. It's also weird to see such a lighthearted naval film with the knowledge of what Hitler was already doing at that time. I have to try to suspend all knowledge to submerge myself into a made up fantasy land.			
5208	5208	1	Yes this a B-grade horror. But at least the producers, directors, and cast does not pretend this flick is manna from heaven. The plot is corny, a psychotic serial killer on his way to execution is splashed with	0		fn

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
		genetic acid turning him into a snow man. The snowman a.k.a. Jack Frost then goes on a murdering rampage to find the small town sheriff that finally arrested him. With a limited budget the crew had to make do with limited special effects, most of the money appears to spent on the snowman's costume. Particullary difficult shots are managed by cartoons or pan away shots (shots where the camera moves away to disguise the details). This is no kid's movie and should not be confused with Disney movie of the same title. If you do not let your children watch pg-13 movies alone than parents should not let their kids watch this movie. This movie has two claims to fame. 1. The beatiful Shannon Elizabeth (American Pie)did her first major movie role. The scene where Jack Frost attacks Shannon Elizabeth is worth watching a few times. 2. This movie has the worst snowman joke ever. The joke is so bad that the directors			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
			credit the joke teller in the credit list. Hard to describe this one -- if you were a fan of Russ Meyer films back in the day, you will surely be pleased to see that Haji is still looking really hot, though Forry Ackerman has not fared so well (what is he doing still making these movies anyway? If I go up to him with a camera will he be in my movie?). It was a pretty fun premise -- a superhero whose giant mammaries are her secret weapon -- but sometimes it did not pan out for the whole length, and the jokes were on a level with your average Joe E. Brown comedy (or, Abbott and Costello if that's your thing) -- basically just bad puns. Still, I found this movie fascinating to watch, and for more than 2 reasons. Good job, but still a fundamentally flimsy production.			
10769	10769	0		1		fp
21063	21063	1	Having enjoyed Joyce's complex novel so keenly I was prepared to be disappointed by Joseph	1		tp

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
		Strick's and Fred Haines's screenplay, given the fabulous complexity of the original text. However, the film turned out to be very well done and a fine translation of the tone, naturalism, and levity of the book. It certainly helps to have read the original text before viewing the film. I imagine the latter would seem disjointed, with very odd episodes apparently randomly stitched together, without a prior reading of the text to help grasp the plot. It's amazing to see how "filthy" the film is, given that it was shot in Dublin in 1967. The Irish film censors only, finally, unbanned it for viewing by general audiences in Ireland as late as 2000 (it was shown to restricted audiences in a private cinema club, the Irish Film Theatre, in the late 1970s). Joyce's eroticism is not simply naturalistic and raunchy, it offers many wildly "perverse" episodes.			

	FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
			Never mind that so many of these fetishes were unacceptable when the book was published in 1922 - they were still utterly taboo when the film was made in 1967. It is astonishing and heartening to watch the cream of the Irish acting profession of the 1960s, respected players all, daring to utter and enact Joyce's hugely transgressive text with such gusto. Bravo!			
22092	22092	1	Spoilers Following: I picked up the book "Evil Angels" when it first came out knowing nothing of the case. Just to give the press and the Austrialian people a break here, I was quite far into it before I began to question the Chamberlain's guilt. The author obviously intended the reader to understand why the public jumped to the conclusions they did. John Bryson told the story just as it was presented to the jurors (and picked up by the press) of the arterial spray, the actelone (??) plates, Dr.	1		tp

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
		James Cameron's certainty that the collar was cut with scissors, that a baby could not be taken whole from her clothes with the buttons still done up, bloody hand print, etc. all quite convincingly. After all, these were experts in their fields who were testifying with no apparent reason to lie, and the fact that the evidence was completely wrong wasn't apparent to me at all. It was also highly technical evidence, difficult for a layman to understand. To this point, beyond some hearsay testimony in the trials, hardly anyone had ever heard of a dingo attacking a human; people didn't believe it was possible. The public was suspicious of the Seventh Day Adventists, whose origins made them appear to be a cult, and all sorts of wild beliefs about them contributed to the appearance of guilt. Were it not for dedicated, selfless lawyers who worked relentlessly to investigate and			

FileIndex	Category	Review	Classifier1Prediction	Classifier1ErrorType	Classifie
		counter the trial testimony, finding Azaria's clothes later would not have been enough to get Lindy out of jail. The book shook me for that reason, and I've been reluctant to come to a conclusion about anyone's guilt ever since (excepting OJ of course). I was thrilled that a movie was going to be made about the case and don't think it could have been done better. I've always liked Sam, who I could identify with completely, and Meryl was perfect as always. Beautiful photography, haunting music. I think it's not only a very good, but a very important, movie. Too bad it didn't receive more publicity at the time it was released.			

In [ ]: