Allen Roberts

DS7337 Natural Language Processing, Fall 2022

Homework 2

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
import re
from nltk.stem.porter import PorterStemmer, wn
from nltk.corpus import stopwords
from gensim.models.word2vec import Word2Vec
from gensim.models import KeyedVectors
from sklearn.cluster.bicluster import SpectralCoclustering
from sklearn.metrics import ConfusionMatrixDisplay
from sklearn.metrics.cluster.bicomatrix_utilities import get_modulation
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
from sklearn.manifold import TSNE
from sklearn.decomposition import PCA
from matplotlib import pyplot as plt
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
from PIL import Image
from urllib.request import urlopen
import sqlite3 as sl
import os
import sys
import joblib
from datetime import datetime
from sklearn.metrics import ConfusionMatrixDisplay
from sklearn.model_selection import GridSearchCV
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier
from sklearn.svm import SVC
from sklearn.metrics import classification_report, ConfusionMatrixDisplay
from sklearn.metrics import roc_curve, auc, RocCurveDisplay
from sklearn.preprocessing import LabelBinarizer
from sklearn.multiclass import OneVsRestClassifier
from itertools import cycle
from scipy.interp import interp
%matplotlib inline
%matplotlib notebook
%load_ext autotime
%load_ext autoreload
%autoreload 2
```

mcu
one
thor
really
movie
chris
thunder
thor
love
pretty

orb
character
peter
groot
guardians
gamora
galaxy
peter quill
nebula
zoe saldana

part
great
deadpool
men
franchise

great
film
movie
deadpool
time

perfect
entertainment
sheer
loud
end
enjoy
bliss
film

funny
feel
kids
school
everything
think

better
good
quality
script
jokes
movie

start
awful
know
watch
fun
movie
trash
maybe

character
movie
one
feel
movie

standard
little
original
least
dumbed

deadpool
guess
adding
first
reference

throughout
better
cast
expected
movie
romance
hard

line
movie
action
great
film
romance
hard

marvel
one
really
huge
bad
fanboy

culture
good
one
large
references

overall
rom
stale
liked
enough
funny
jokes

reality
comedy
good
enjoy
movie

character
much
almost
plot
grandiosity
twist

sequences
bland
hollow
jokes
string
fight
gag

film
scene
onwards
opening
time
incoherent

believe
misfits
mcu
movie
called

okay
stupidity
parts
potential
premise
overloaded

squad
guardians
movie
suicide
story
funny

gunn
galaxy
movie
guardians

studded
truth
actually
see
people

first
new
part
abomination

deadpool
movie
character

spending
everything
point
time
fine

funny
write
fast
sure
story
attempt
entertainment

joke
leaves
everything
check
even

sum
typing
reduced
bothered
essay
thor

audience
really
superheroes
film
movie
end

lazy
big
mediocre
try
carry
budget
effort
precipice

funniest
heroes
excellent
great
especially
film
movie

whole
chair
end
plastic
instead
cut

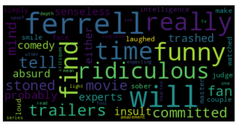
marvel
good
disney
think
bad
want

russell
crowe
christian
bale
movie
great
good

movie
pure
comedy
good
jokes
film
rated

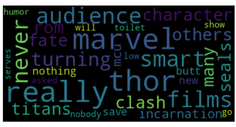
bad
humour
parody
min
ridiculous

Cluster: 45
IMDB Review Links
https://www.imdb.com/review/rw6923557/
IMDB Review Texts



Cluster: 46
IMDB Review Links
https://www.imdb.com/review/rw4161726/
IMDB Review Texts

Cluster: 47
IMDB Review Links
https://www.imdb.com/review/rw8315634/
IMDB Review Texts

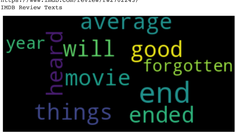


Cluster: 48
IMDB Review Links
https://www.imdb.com/review/rw7293764/
https://www.imdb.com/review/rw4297422/
https://www.imdb.com/review/rw6991276/
https://www.imdb.com/review/rw4379195/
https://www.imdb.com/review/rw4855093/
https://www.imdb.com/review/rw4229571/
https://www.imdb.com/review/rw3415632/
https://www.imdb.com/review/rw3708156/
https://www.imdb.com/review/rw2864456/
https://www.imdb.com/review/rw2076017/
https://www.imdb.com/review/rw2036741/
https://www.imdb.com/review/rw7339951/
https://www.imdb.com/review/rw7558618/
https://www.imdb.com/review/rw7248676/
https://www.imdb.com/review/rw7224587/
IMDB Review Texts

Cluster: 49
IMDB Review Links
https://www.imdb.com/review/rw8312037/
IMDB Review Texts

Cluster: 50
IMDB Review Links
https://www.imdb.com/review/rw4009143/
IMDB Review Texts

Cluster: 51
IMDB Review Links
https://www.imdb.com/review/rw2702243/
IMDB Review Texts

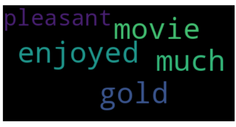


Cluster: 52
IMDB Review Links
https://www.imdb.com/review/rw8093318/
IMDB Review Texts

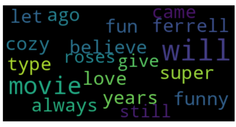
Cluster: 53
IMDB Review Links
https://www.imdb.com/review/rw6124126/
IMDB Review Texts

Cluster: 54
IMDB Review Links
https://www.imdb.com/review/rw8138680/
IMDB Review Texts

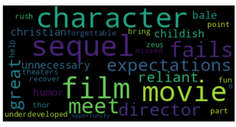
Cluster: 55
IMDB Review Links
https://www.imdb.com/review/rw7999963/
IMDB Review Texts



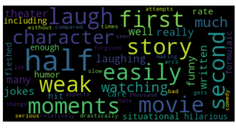
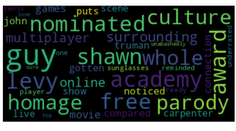
Cluster: 56
IMDB Review Links
https://www.imdb.com/review/rw3419493/
IMDB Review Texts



Cluster: 57
IMDB Review Links
https://www.imdb.com/review/rw8329077/
IMDB Review Texts



Cluster: 58
IMDB Review Links
https://www.imdb.com/review/rw8311274/
IMDB Review Texts



Cluster: 59
IMDB Review Links
https://www.imdb.com/review/rw2079109/
IMDB Review Texts