

Analyzing Movie Reviews

Natural Language Processing



Background

Why are movie reviews relevant?

Background

Why are movie reviews relevant?

- Commonly used by consumers to decide if a movie is worth the price and time.

Background

Why are movie reviews relevant?

- Commonly used by consumers to decide if a movie is worth the price and time.
- Influence a global audience.

Background

Why are movie reviews relevant?

- Commonly used by consumers to decide if a movie is worth the price and time.
- Influence a global audience.
- Have become a crucial aspect of the movie industry

Goal

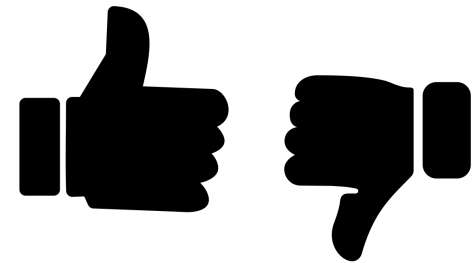


- **Gaining insights** from online movie reviews to inform Pinewood Studio 2021 strategic planning.

Goal



- **Gaining insights** from online movie reviews to inform Pinewood Studio 2021 strategic planing.
- Specifically, we want to answer the following **questions**:
 - What aspects of a movie people tend to consider relevant?
 - When people like or dislike a movie, what aspects of the movie do they like or dislike?



Methodology

Natural Language Processing

- **Data:**

A random sample of 25,000 IMDB movie reviews from (Kaggle Dataset).
The raw data has two columns (review, sentiment), and 50,000 rows.



Stanford University

kaggle

Methodology

Natural Language Processing

- **Data:**

A random sample of 25,000 IMDB movie reviews from (Kaggle Dataset).
The raw data has two columns (review, sentiment), and 50,000 rows.



Stanford University

kaggle

- **Tools:**

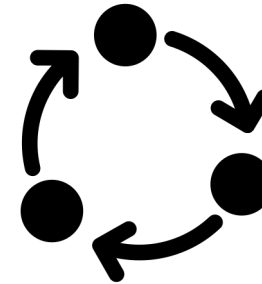
- Spacy
- NLTK
- Scikit-learn (NMF, TF-IDF)



Methodology

Natural Language Processing

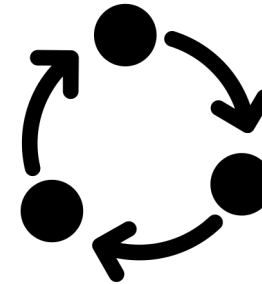
- Data preprocessing:
 - removed punctuations, and stopwords
 - included only nouns, verbs, and adjectives
 - lemmatization



Methodology

Natural Language Processing

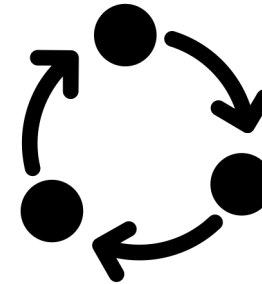
- **Data preprocessing:**
 - removed punctuations, and stopwords
 - included only nouns, verbs, and adjectives
 - lemmatized
- **Data vectorization:**
 - Term Frequency-Inverse (tf-idf)



Methodology

Natural Language Processing

- **Data preprocessing:**
 - removed punctuations, and stopwords
 - included only nouns, verbs, and adjectives
 - lemmatized
- **Data vectorization:**
 - Term Frequency-Inverse (tf-idf)
- **Dimensionality Reduction and Topic Modeling:**
 - Non-Negative Matrix Factorization



10
topics



10
topics

- **Horror**

10 topics

- **Horror**
- **Bad Movie**

10 topics

- **Horror**
- **Bad Movie**
- **Good Movie**

10 topics

- **Horror**
- **Bad Movie**
- **Good Movie**
- **Comedy**

10 topics

- **Horror**
- **Bad Movie**
- **Good Movie**
- **Comedy**
- **Romance**

10 topics

- **Horror**
- **Bad Movie**
- **Good Movie**
- **Comedy**
- **Romance**
- **Tv Show**

10 topics

- **Horror**
- **Bad Movie**
- **Good Movie**
- **Comedy**
- **Romance**
- **Tv Show**
- **Characters & Plot**

10 topics

- **Horror**
- **Bad Movie**
- **Good Movie**
- **Comedy**
- **Romance**
- **Tv Show**
- **Characters & Plot**
- **Acting Performance**

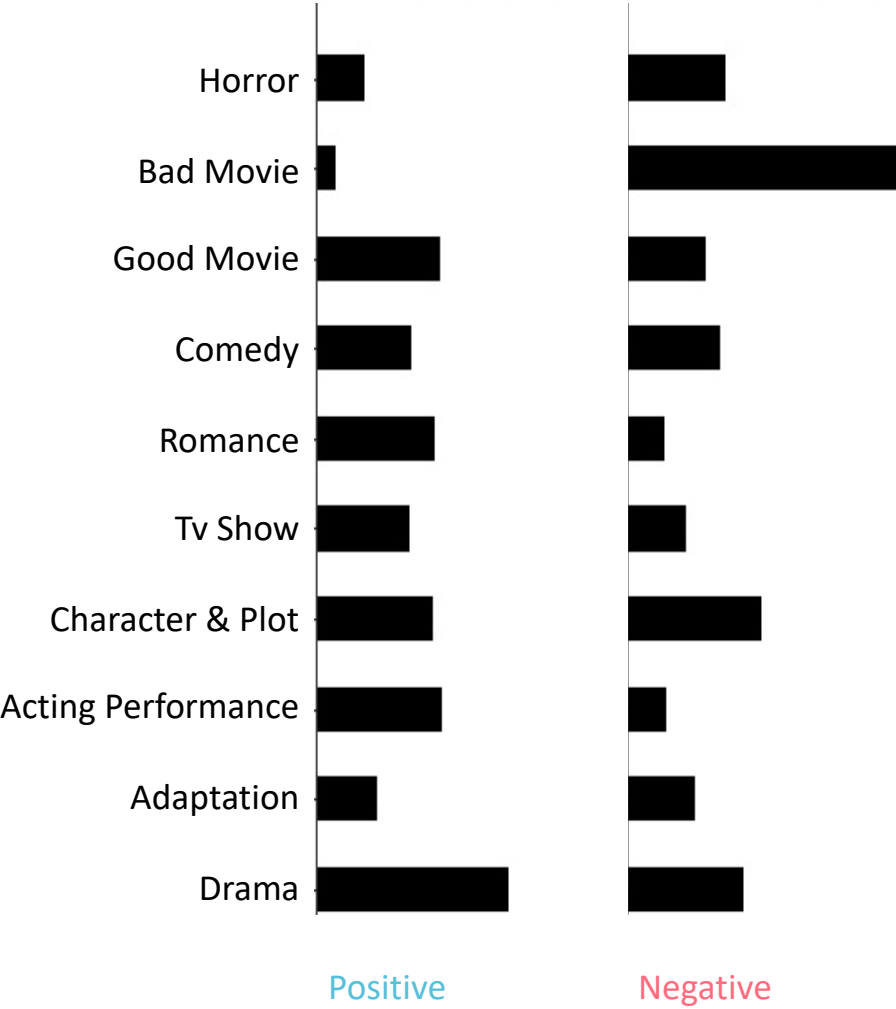
10 topics

- **Horror**
- **Bad Movie**
- **Good Movie**
- **Comedy**
- **Romance**
- **Tv Show**
- **Characters & Plot**
- **Acting Performance**
- **Adaptation**

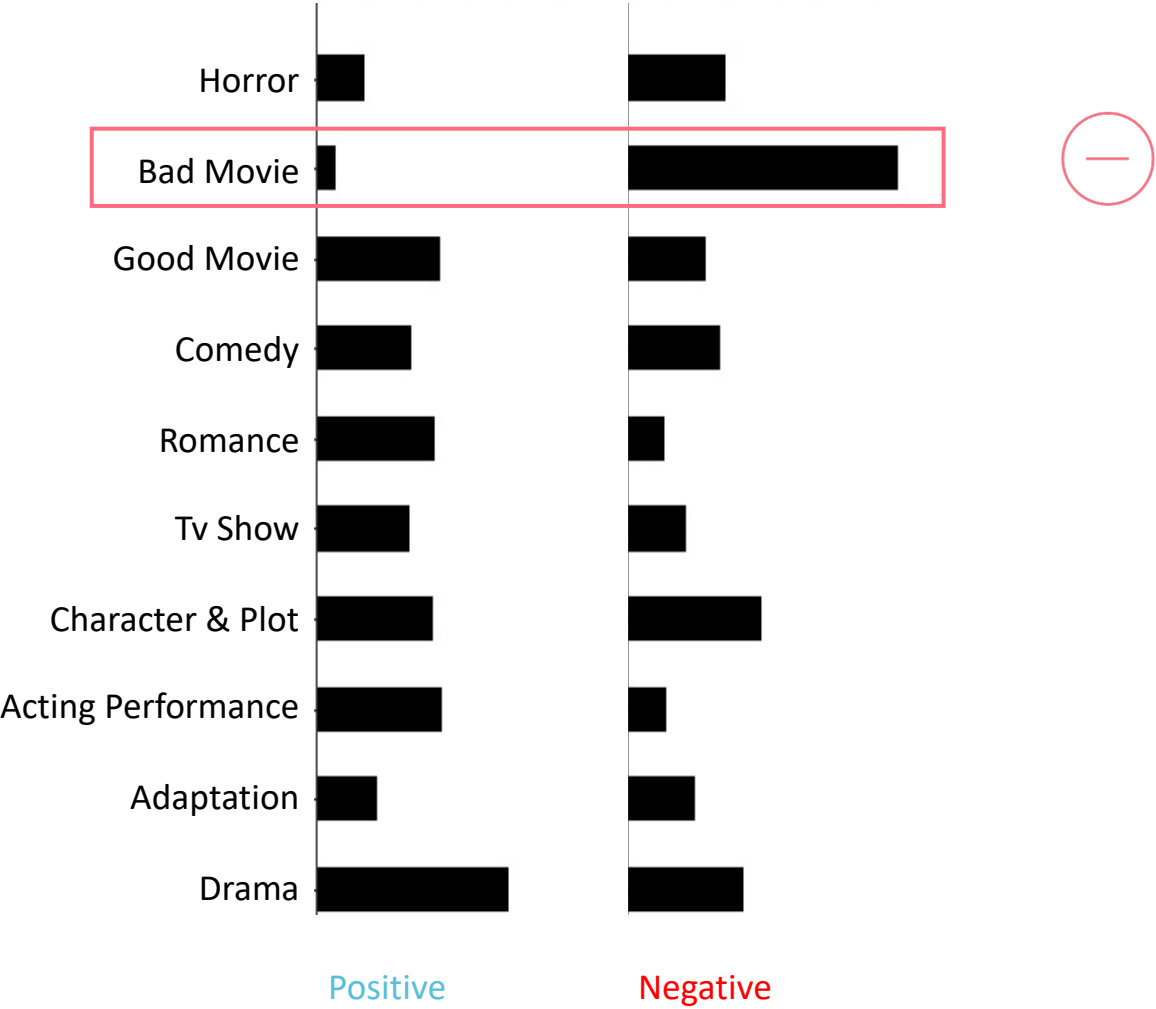
10 topics

- **Horror**
- **Bad Movie**
- **Good Movie**
- **Comedy**
- **Romance**
- **Tv Show**
- **Characters & Plot**
- **Acting Performance**
- **Adaptation**
- **Drama**

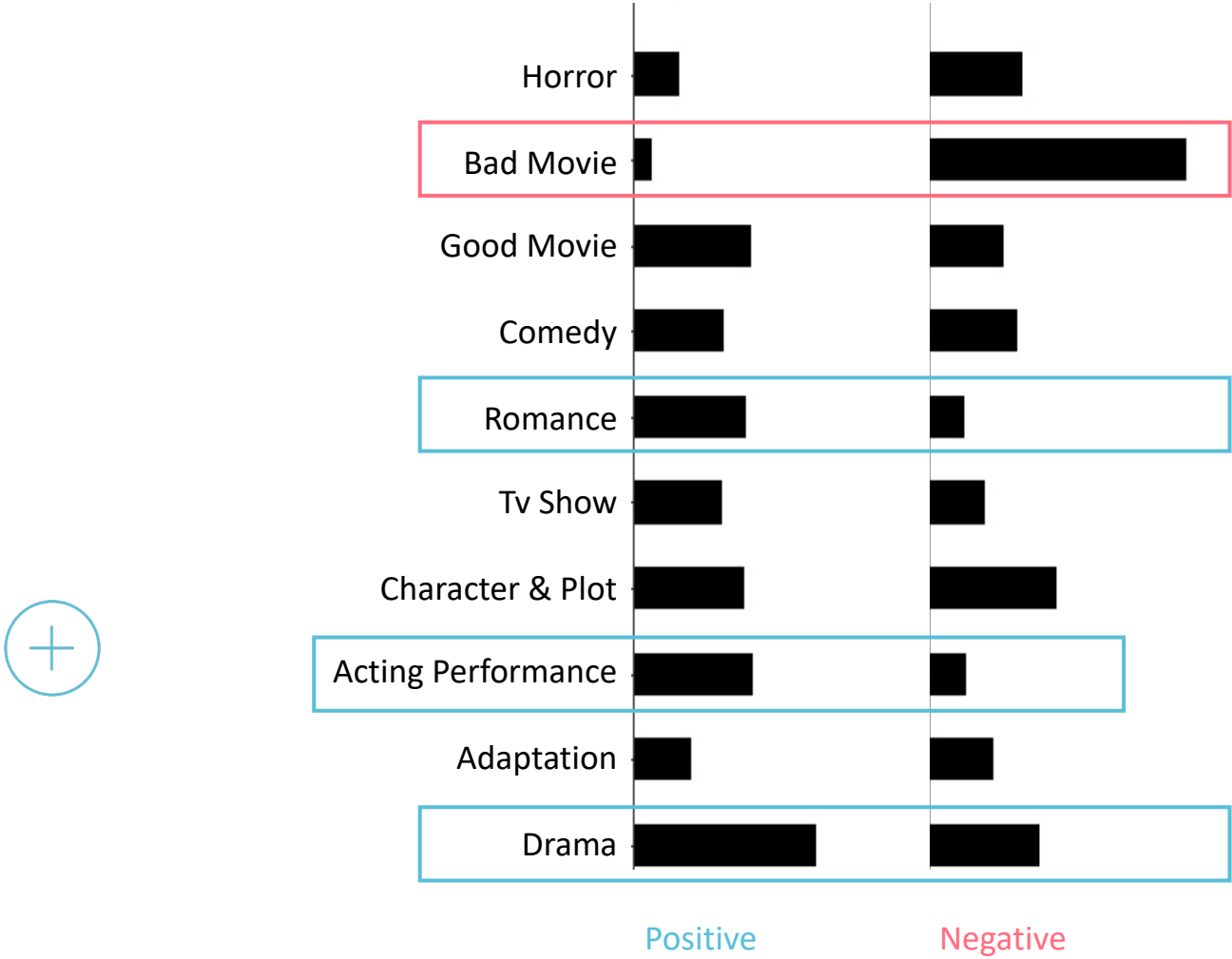
Proportion of topics in **positive** reviews versus **negative** reviews



Proportion of topics in **positive** reviews versus **negative** reviews



Proportion of topics in **positive** reviews versus **negative** reviews

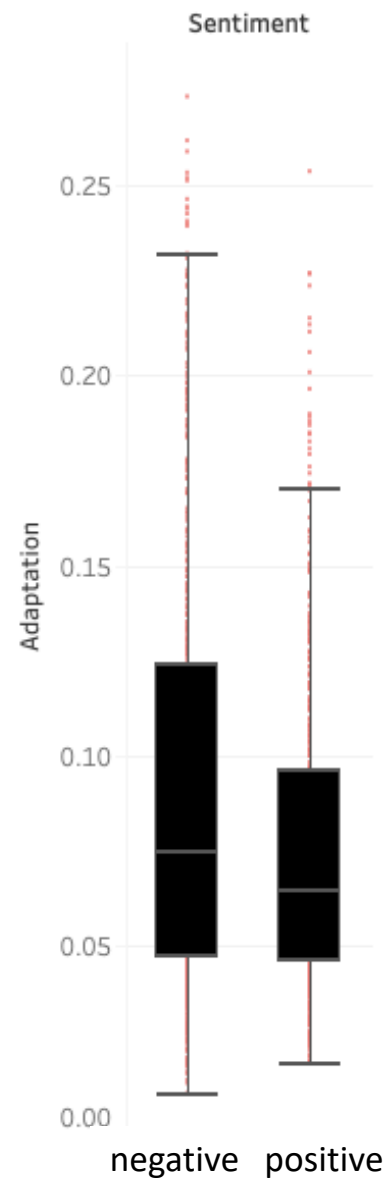


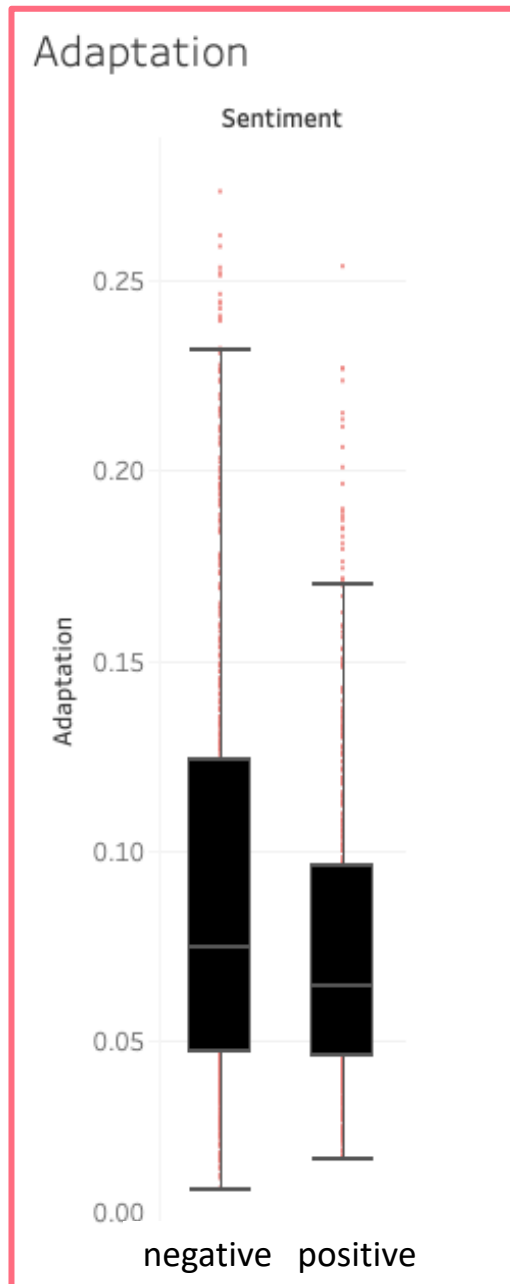
“

When this movie came out, I had seen Geena Davis play only soft, feminine roles... This movie was anything but soft and feminine. Great lines, great action...she and Samuel L really clicked. Too violent for the kiddies, but if you and your significant other are trying to agree on a movie, try this one on for size. Go Geena! Go Geena!

”

Adaptation





“

After reading the book, Heart of Darkness, the movie did not do it justice. The movie puts the book to shame and anyone who has not experienced the book would frown upon the story and plot because it was portrayed so poorly by the movie. In the film, the characters and set were just some of the let downs that occurred in the movie. The director left out so many important and interesting aspects of the book that made it one of the best literary works ever made. Of course any book is better than the movie but these weren't even comparable. Joseph Conrad as a writer was brilliant in vocabulary and the cleverness of the written word. The movie doesn't even start to show any of this. Some of the very important and influential scenes from the book were completely left out.

”

Words & Topics

Horror

horror (1.49)
plot (0.89)
kill (0.60)
killer (0.55)
shoot (0.48)
fight (0.45)
low (0.44)
monster (0.39)
blood (0.39)
scary (0.38)
dead (0.37)
cut (0.36)
night (0.36)
death (0.35)



Comedy

funny (2.57)

comedy (1.77)

laugh (1.43)

joke (0.82)

humor (0.52)

hilarious (0.42)

fun (0.34)

stupid (0.25)

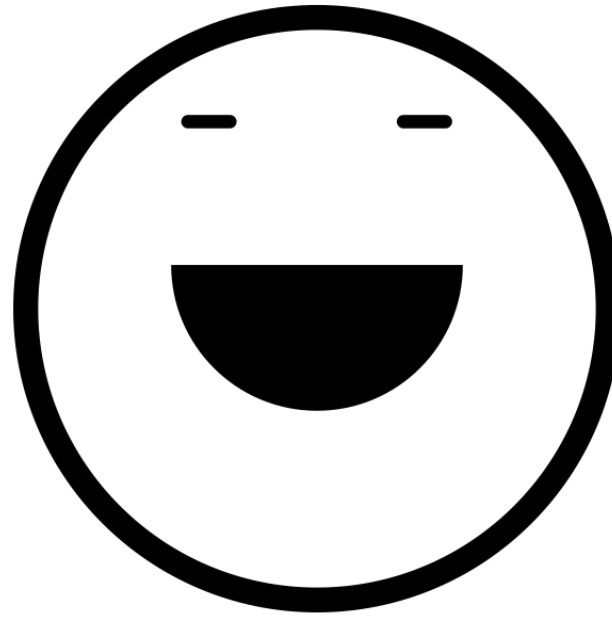
enjoy (0.22)

romantic (0.154)

loud (0.14)

silly (0.36)

parody (0.36)



Comedy

funny (2.57)

comedy (1.77)

laugh (1.43)

joke (0.82)

humor (0.52)

hilarious (0.42)

fun (0.34)

stupid (0.25)

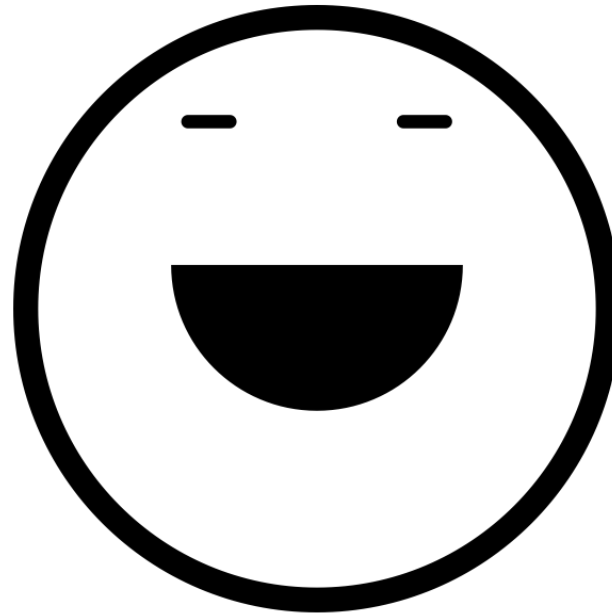
enjoy (0.22)

romantic (0.154)

loud (0.14)

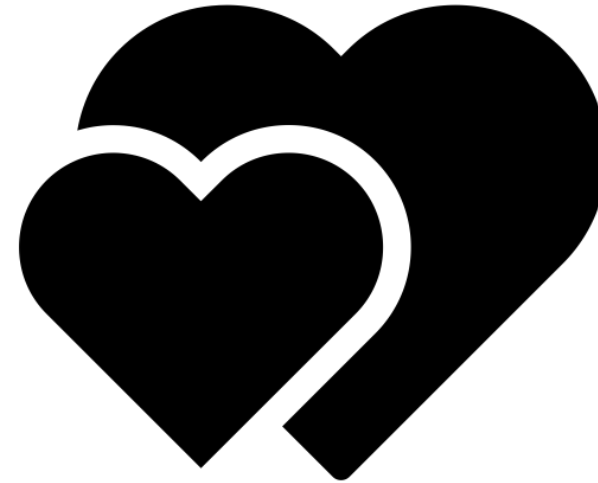
silly (0.36)

parody (0.36)



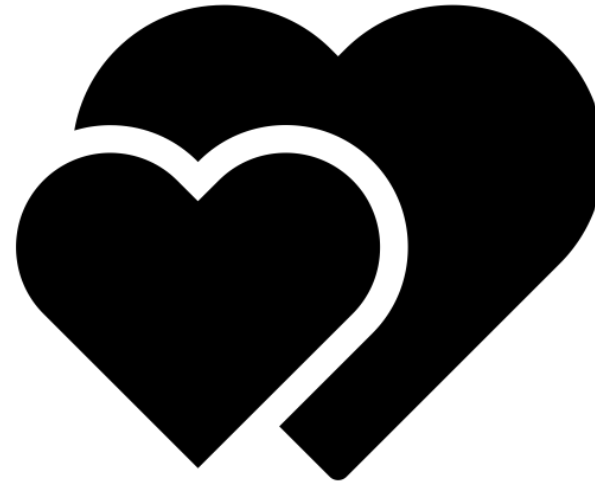
Romance

love (3.68)
story (0.411)
song (0.410)
music (0.38)
enjoy (0.28)
beautiful (0.27)
wonderful (0.25)
favorite (0.23)
romantic (0.15)
scene (0.22)
romance (0.18)
dance (0.18)
heart(0.17)
amazing (0.16)
friend (0.16)
true (0.16)



Romance

love (3.68)
story (0.411)
song (0.410)
music (0.38)
enjoy (0.28)
beautiful (0.27)
wonderful (0.25)
favorite (0.23)
romantic (0.15)
scene (0.22)
romance (0.18)
dance (0.18)
heart(0.17)
amazing (0.16)
friend (0.16)
true (0.16)



Romance

love (3.68)

story (0.411)

song (0.410)

music (0.38)

enjoy (0.28)

beautiful (0.27)

wonderful (0.25)

favorite (0.23)

romantic (0.15)

scene (0.22)

romance (0.18)

dance (0.18)

heart(0.17)

amazing (0.16)

friend (0.16)

true (0.16)



TV Show

episode (2.22)

series (2.07)

tv (0.84)

season (0.71)

television (0.28)

dvd (0.20)

scifi (0.15)

cartoon (0.13)

writer (0.10)

alien (0.10)

pilot (0.10)

planet (0.09)

long (0.09)

program (0.092)

animation(0.090)



Characters & Plot

character (2.72)
story (2.08)
main (0.52)
interesting (0.45)
plot (0.42)
end (0.36)
development (0.27)
real (0.25)
true (0.18)
dialogue (0.17)
actor (0.17)
depth (0.16)
believable (0.15)
twist (0.14)
weak(0.14)



Characters & Plot

character (2.72)
story (2.08)
main (0.52)
interesting (0.45)
plot (0.42)
end (0.36)
development (0.27)
real (0.25)
true (0.18)
dialogue (0.17)
actor (0.17)
depth (0.16)
believable (0.15)
twist (0.14)
weak(0.14)



Characters & Plot

character (2.72)

story (2.08)

main (0.52)

interesting (0.45)

plot (0.42)

end (0.36)

development (0.27)

real (0.25)

true (0.18)

dialogue (0.17)

actor (0.17)

depth (0.16)

believable (0.15)

twist (0.14)

weak(0.14)



Adaptation

book (2.65)

read (1.11)

novel (0.52)

adaptation (0.30)

write (0.21)

story (0.20)

dissapointed (0.27)

screenplay (0.25)

true (0.18)

disappoint (0.17)

depth (0.16)

believe (0.15)



Adaptation

book (2.65)

read (1.11)

novel (0.52)

adaptation (0.30)

write (0.21)

story (0.20)

dissapointed (0.27)

screenplay (0.25)

true (0.18)

disappoint (0.17)

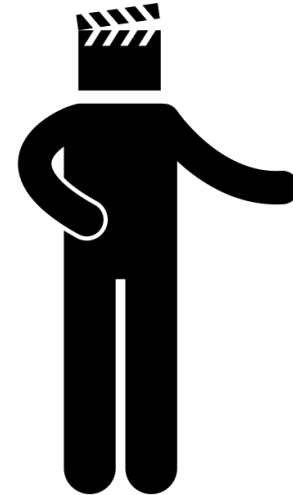
depth (0.16)

believe (0.15)



Acting Performance

great (2.65)
actor (1.36)
performance (1.21)
role (0.30)
cast (0.89)
excellent (0.53)
star (0.64)
wonderful (0.40)
actress (0.35)
song (0.28)
perfect (0.26)
support (0.26)
talent (0.26)
brilliant (0.25)
fantastic (0.23)
direction (0.20)



Drama

life (1.38)
man (1.21)
woman (0.87)
young (0.70)
world (0.68)
family (0.61)
live (0.59)
war (0.55)
child (0.48)
father (0.45)
wife (0.40)
work (0.38)
friend (0.38)
mother (0.36)
son (0.33)
human (0.30)
daughter (0.28)



Good Movie

good (3.75)
acting (0.49)
enjoy (0.21)
recommend (0.20)
nice (0.20)
fun (0.19)
excellent (0.13)
worth (0.13)
interesting (0.10)
cool (0.10)
ending (0.10)
actor (0.07)
oscar (0.07)
entertaining (0.07)
actress (0.07)
cast (0.07)
twist (0.07)
plot (0.07)



Good Movie

good (3.75)
acting (0.49)
enjoy (0.21)
recommend (0.20)
nice (0.20)
fun (0.19)
excellent (0.13)
worth (0.13)
interesting (0.10)
cool (0.10)
ending (0.10)
actor (0.07)
oscar (0.07)
entertaining (0.07)
actress (0.07)
cast (0.07)
twist (0.07)
plot (0.07)



Good Movie

good (3.75)

acting (0.49)

enjoy (0.21)

recommend (0.20)

nice (0.20)

fun (0.19)

excellent (0.13)

worth (0.13)

interesting (0.10)

cool (0.10)

ending (0.10)

actor (0.07)

oscar (0.07)

entertaining (0.07)

actress (0.07)

cast (0.07)

twist (0.07)

plot (0.07)



Good Movie

good (3.75)

acting (0.49)

enjoy (0.21)

recommend (0.20)

nice (0.20)

fun (0.19)

excellent (0.13)

worth (0.13)

interesting (0.10)

cool (0.10)

ending (0.10)

actor (0.07)

oscar (0.07)

entertaining (0.07)

actress (0.07)

cast (0.07)

twist (0.07)

plot (0.07)



Bad Movie

bad (3.40)

waste (0.53)

acting (0.053)

awful (0.41)

terrible (0.40)

horrible (0.37)

script (0.36)

actor (0.36)

plot (0.32)

stupid(0.29)

crap (0.29)

garbage (0.13)

ridiculous (0.13)

pathetic (0.13)

god (0.12)

trash (0.12)

end (0.12)

boring (0.12)

dialogue (0.11)

cheap (0.11)



Takeaways

Why the audience liked the movie?

- They liked the plot (sequence of events).
- The movie is not predictable, has an interesting twist.
- Liked the performance of the actors (believable).

Takeaways

- Invest more in the soundtrack when producing a romantic movie.
- Topic modeling on online book reviews is advised to inform script writing of a film adaptation.

Future work

- Additional topic modeling including the whole dataset of 50,000 reviews.

Thank You!

laura.urdapilleta@gmail.com

