

Movie Rating Predictor and Recommender System

Jen Zieger



SCENARIO



Situation

A new streaming service's production company has recognized this issue, and they are looking for a way to recommend movies to viewers differently than some of their competitors.

They are also looking for a new way to predict if their new movie releases will be rated as Fresh or Rotten by audience viewers on Rotten Tomatoes.



Problem Statement

Given movie review information, what features predict if an audience will rate a movie as Fresh or Rotten on Rotten Tomatoes, and how can that information be used to build a recommender system that can be used to exceed user expectations?

STRATEGY

01

DATA
COLLECTION
& CLEANING

02

EXPLORATORY
DATA
ANALYSIS

03

MODELING
& EVALUATION

04

CONCLUSION &
RECOMMENDATIONS

01

DATA COLLECTION & CLEANING





One of the world's most trusted recommendation resources for entertainment – provides fans with a comprehensive guide to what's "Rotten" and "Fresh."

- Ratings:
 - Tomatometer: 0-100%
 - Audience score: 0-5
- Parent Company: Fandango Media, LLC
- 65M+ Monthly Visitors
- Kaggle Dataset: 1.1M critic reviews with 17K movie titles



One of the world's most popular sources for movie, TV, and celebrity content – allows fans to explore movies and shows, and decide what to watch

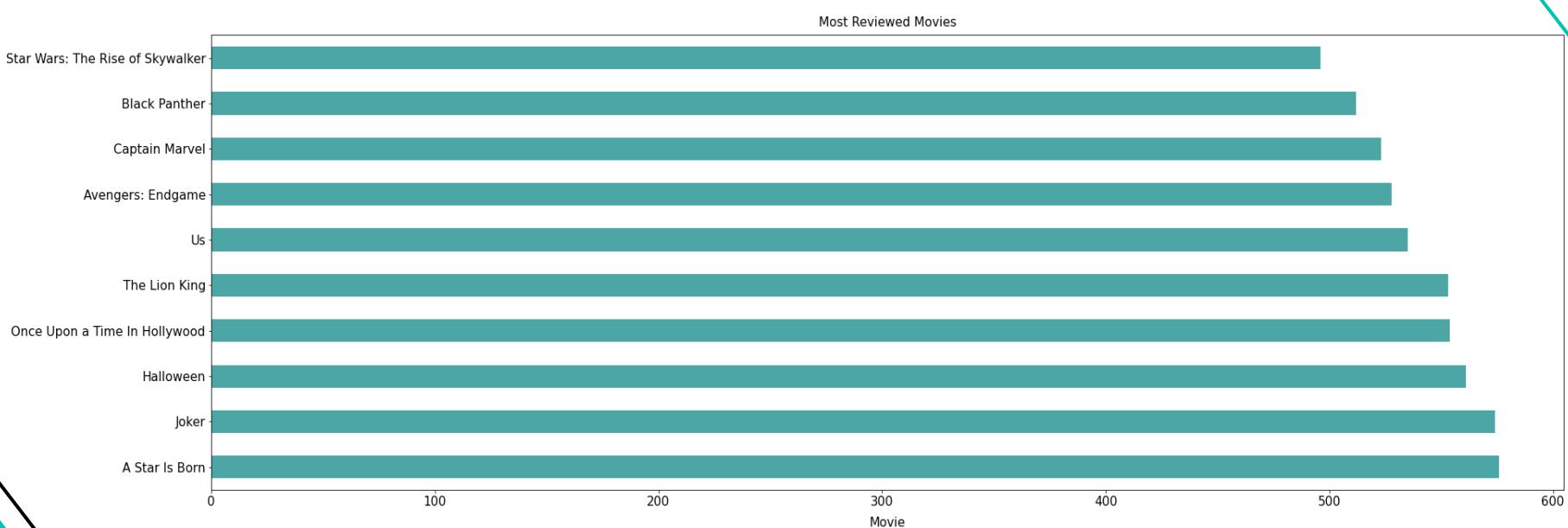
- Ratings:
 - IMDb User: 1-10
- Parent Company: Amazon
- 500M+ Monthly Visitors
- Kaggle Dataset: 85K movie titles and ratings



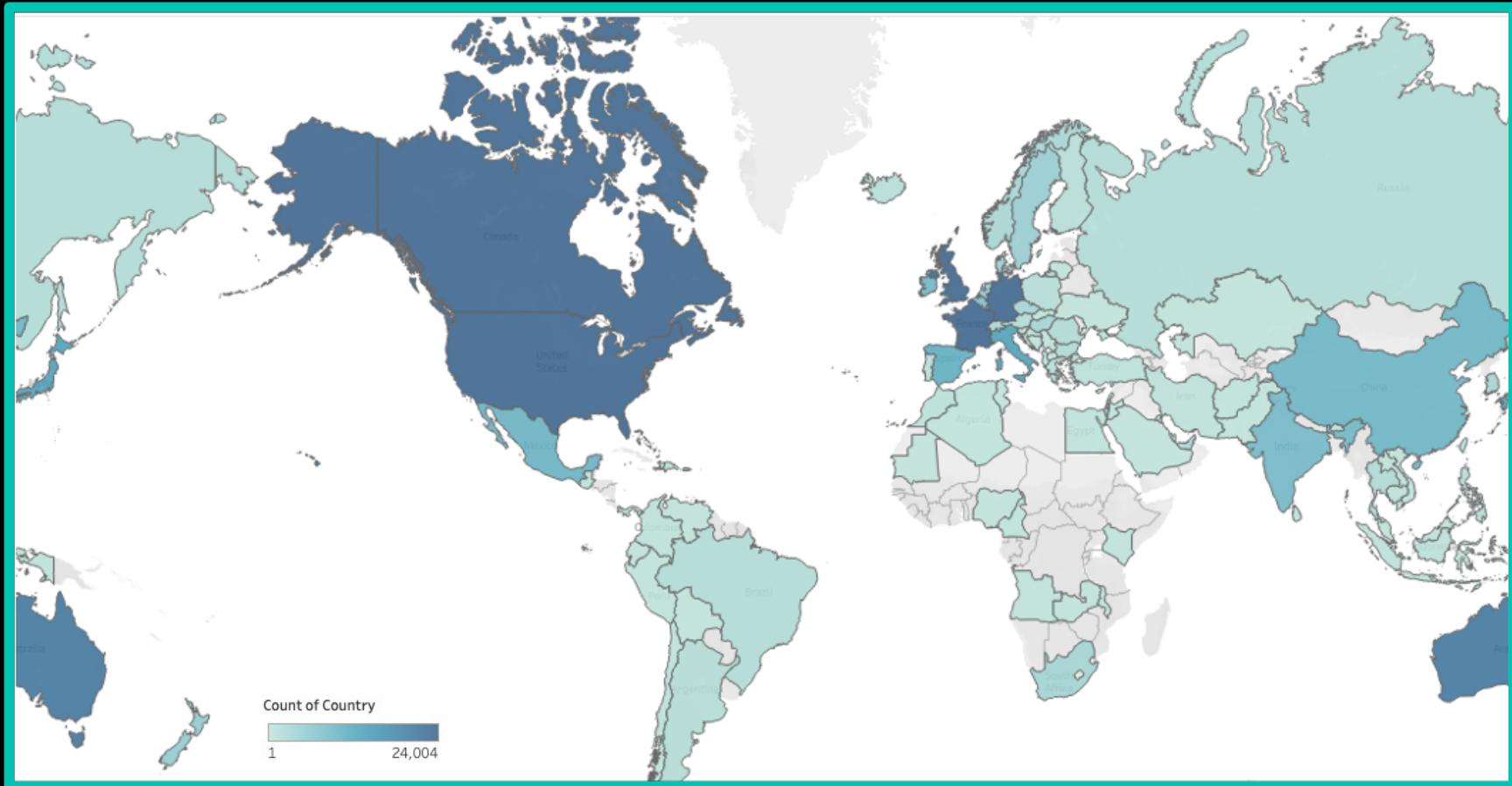
02

EXPLORATORY DATA ANALYSIS

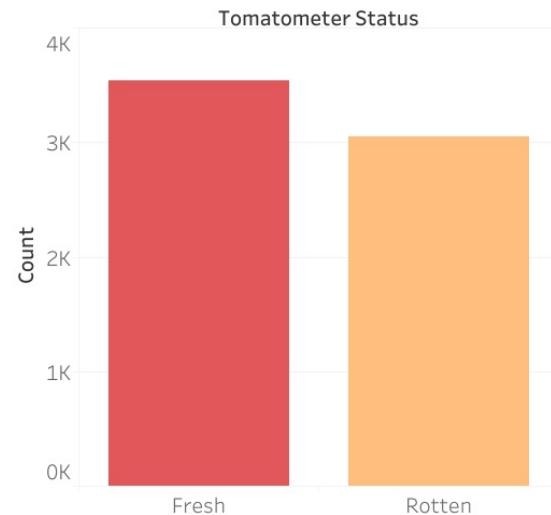
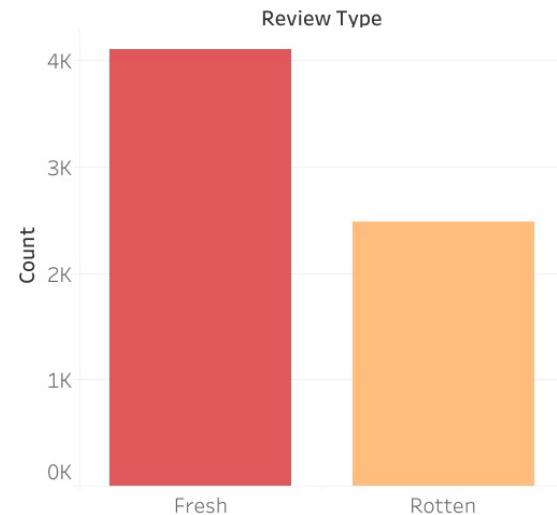
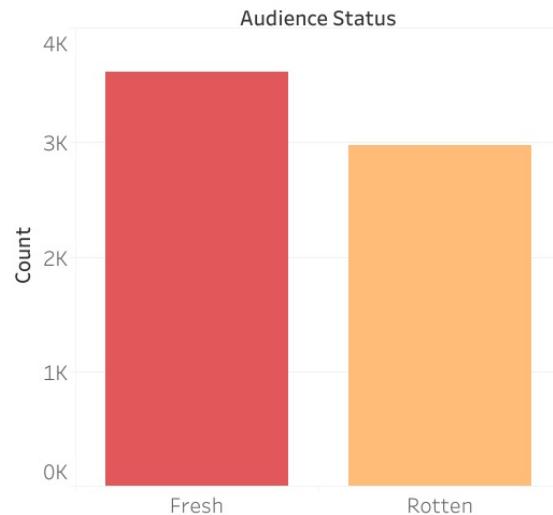
MOST REVIEWED MOVIES



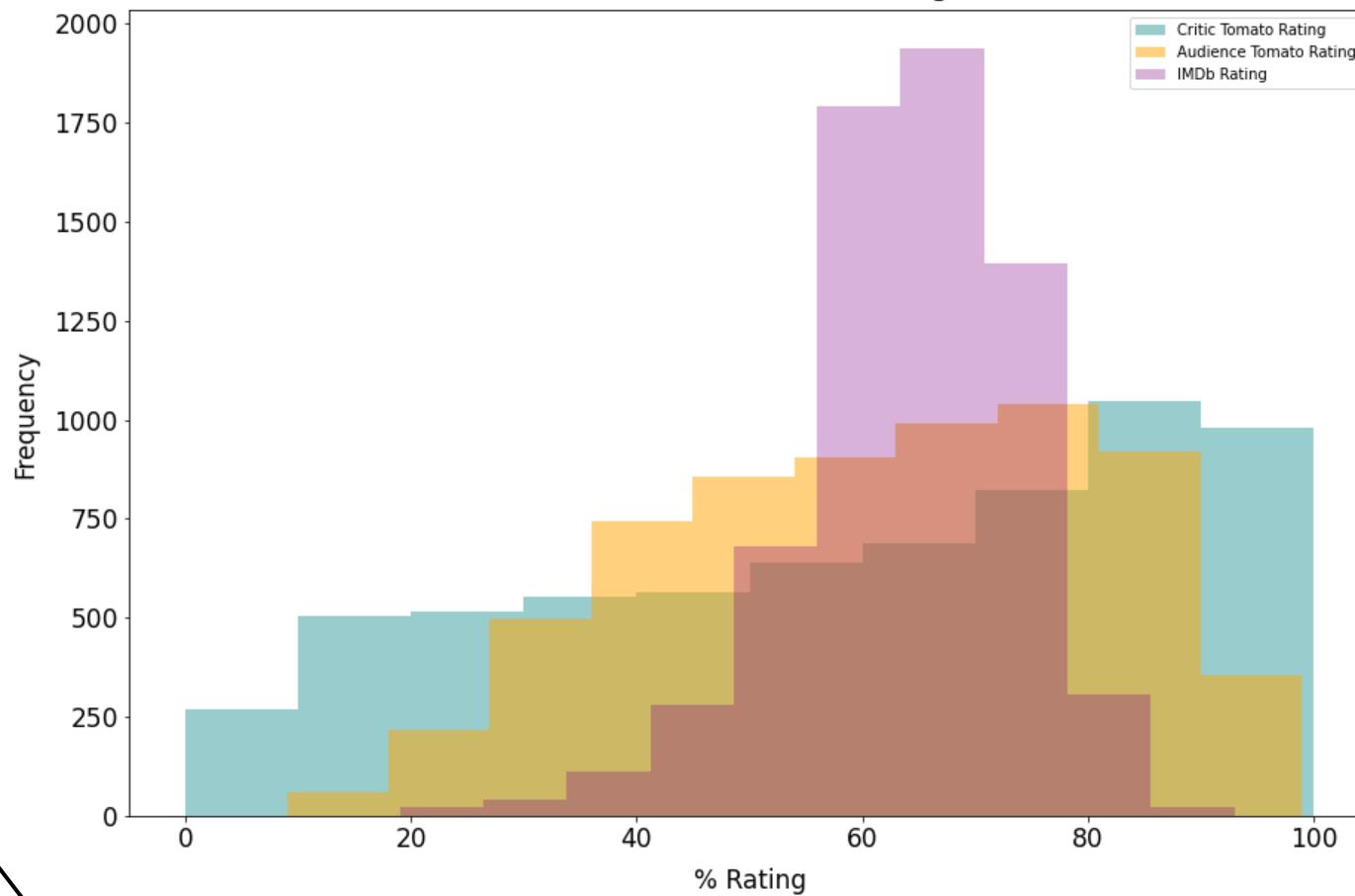
TOP COUNTRIES BY MOVIE COUNT



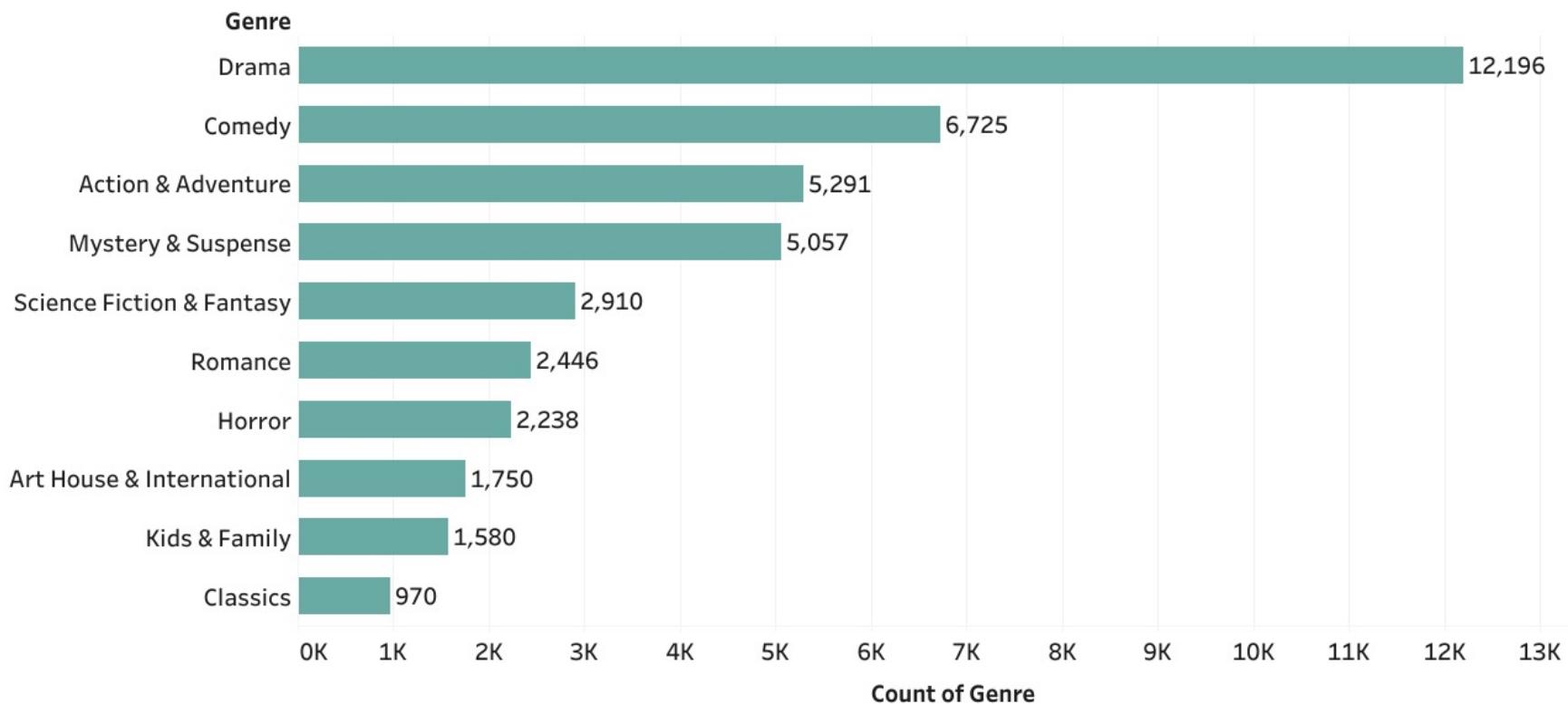
ROTTEN VS FRESH

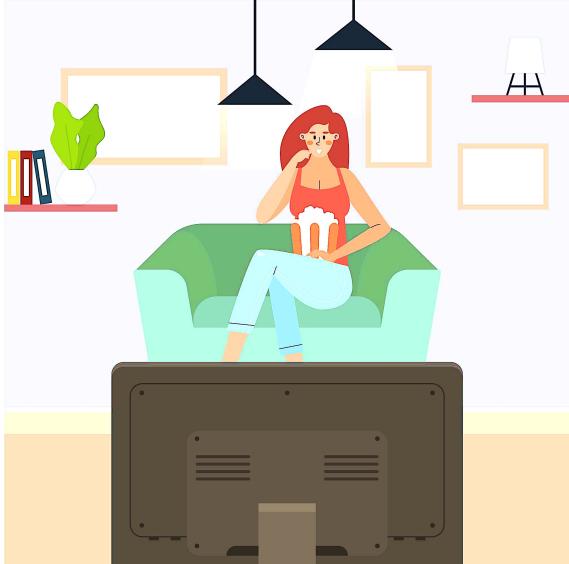


Distribution of Movie Ratings



TOP GENRES





03

MODELING & EVALUATION

PRE-PROCESSING

LOWERCASE

REMOVE CONTRACTIONS

REMOVE SPECIAL CHARACTERS

STOP WORDS

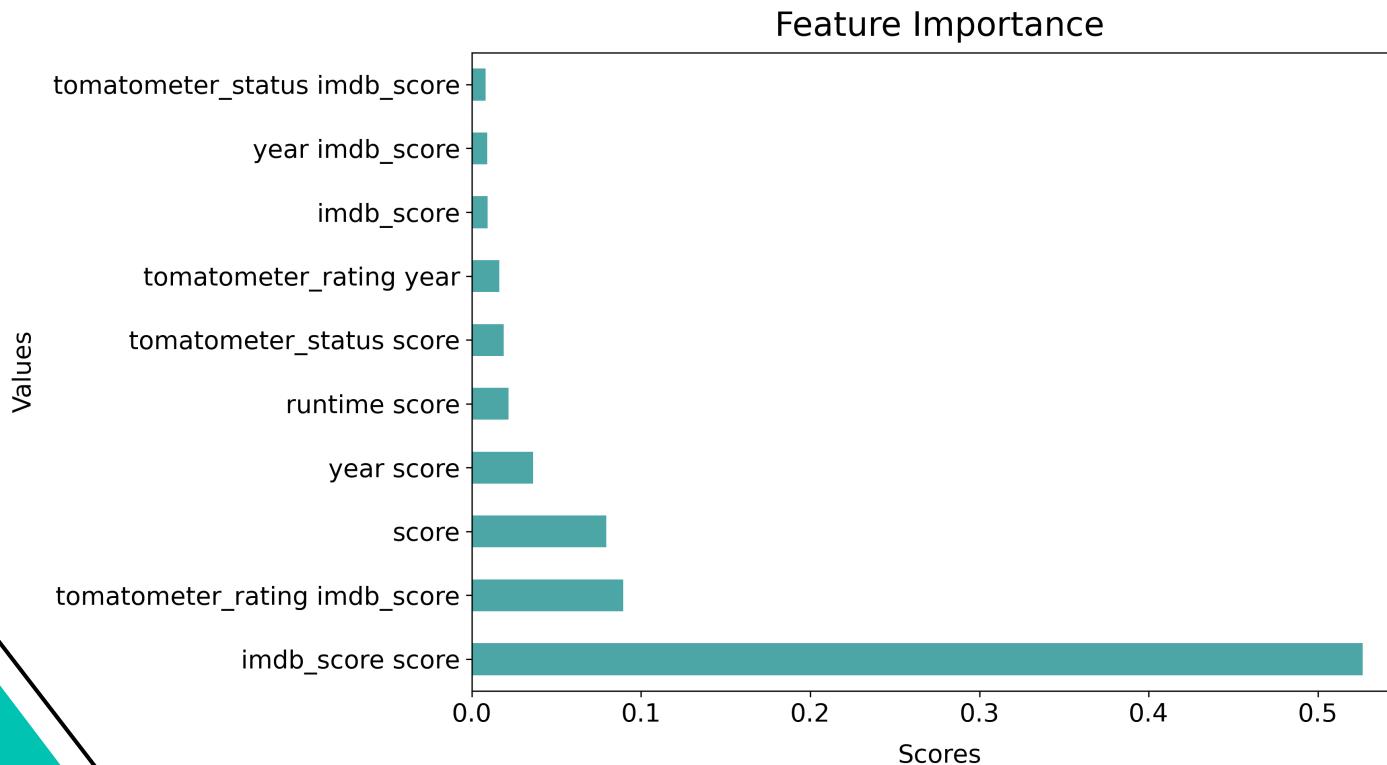
LEMMATIZE

TOKENIZE

TOP 3 CLASSIFICATION MODELS

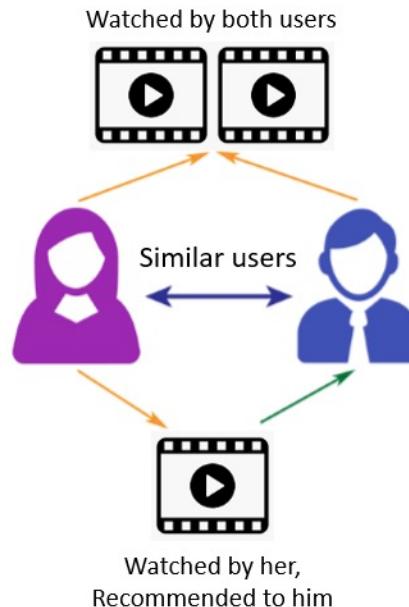
Model	Train Accuracy	Test Accuracy	Precision	Recall	Specificity	Misclass	F1 Score
Gradient Boost	1.00	0.928	0.942	0.925	0.931	0.718	0.933
Bagging	1.00	0.919	0.933	0.918	0.920	0.080	0.926
Decision Tree	0.932	0.880	0.897	0.897	0.877	0.119	0.889

Feature Importance

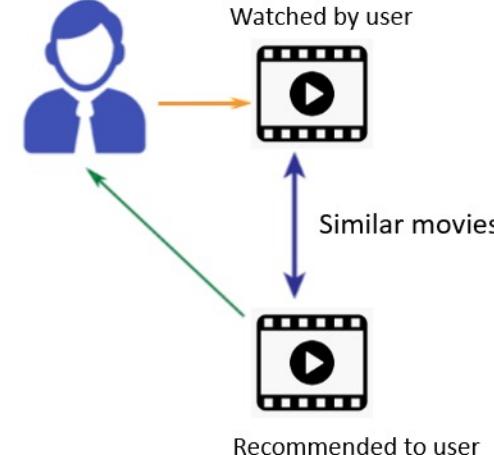


COLLABORATIVE VS CONTENT BASED FILTERING

Collaborative Filtering



Content-Based Filtering



RECOMMENDER SYSTEM MODELING



Collaborative

Recommendations based
on critic's ratings



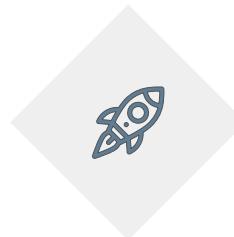
Content Based

Recommendations
based on keyword
filtering



Transformers & Libraries

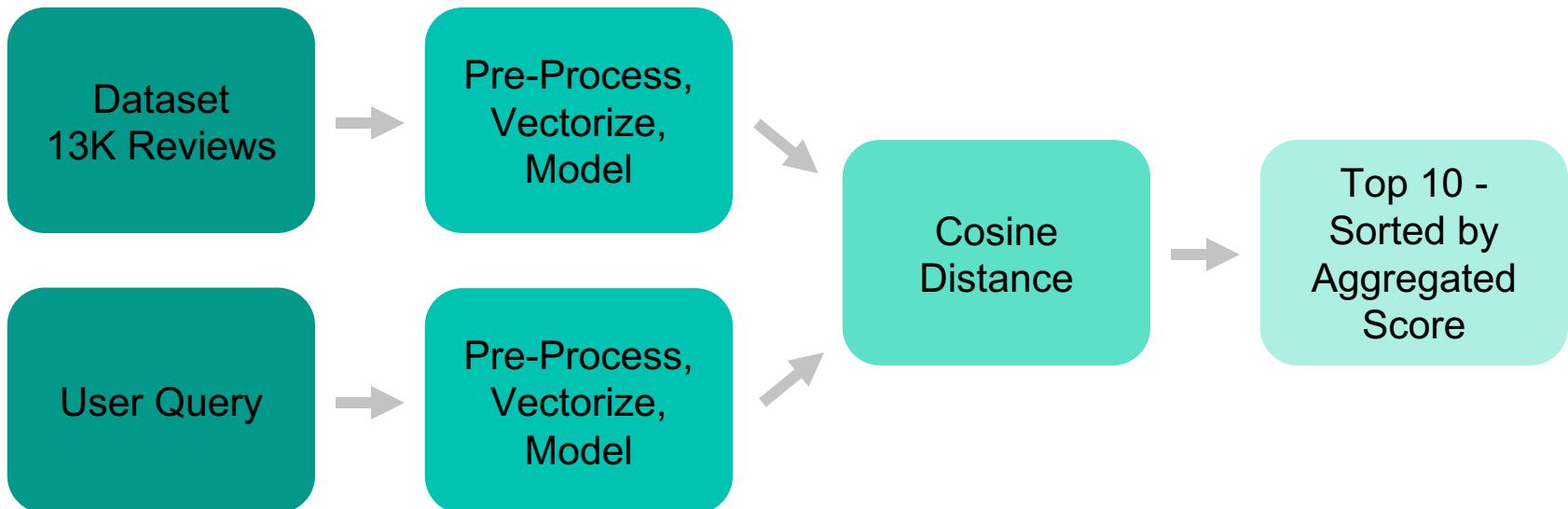
TF-IDF, Count Vectorizer,
Word2Vec, spaCy



Metric

Cosine
Euclidean
Manhattan

FINAL MODEL – CONTENT BASED



05

CONCLUSION & RECOMMENDATIONS



CONCLUSIONS



Classification Model

Gradient Boost Model – 93% Accuracy

Highest Feature Importance:
Score & IMDb Score

Recommender System

Content Based Filtering with TF-IDF
allows users to input a sentence about
what they want to watch and receive
recommendations

RECOMMENDATIONS

Recommender System

Model:

- Test other models and fine-tune existing
- Get ratings and reviews/text from other sources
- Use audience individual Scores
- Try other recommender system models/fine-tune existing
- Add TV Shows

Website:

- Add usernames and login capabilities
- Use analytics tools for traffic information
- Hire a website developer

Classification Model

- Test other models and fine-tune existing
- Get current data
- Model larger set of data
- Add TV Shows
- Predict other types of ratings:
 - Critic, IMDb, etc.
- Use other features:
 - Other rating datasets – Metacritic, Google, etc.
 - Awards, individual audience reviews

DEMO

MovieRec Website

Search Examples

I'm in the mood for
a gut-wrenching
dramatic movie

Schindler's List
The Wrestler
127 Hours
The Attack
Without Limits

College or High
School Comedy

Dazed and Confused
Election
The Spectacular Now
Grosse Pointe Blank
Easy A

Scorsese

Taxi Driver
Raging Bull
The Departed
Mean Streets
The King of Comedy

MovieRecs

Grab some popcorn!
You're about to watch a movie you're actually in the mood for!

Please use the below form to input what you think you'd want to watch.
This can be a title, an actor, a genre, a mood, you name it!

The more detailed you are, the better the recommendations will be.

Some examples include:

- * I want to watch a scary horror movie
 - * Scorsese
- * I want to watch a college or high school comedy
- * I'm in the mood for a serious movie with lots of drama
- * I want to watch a movie with Jordan Peele in it
 - * International
- * Meryl Streep Comedy
 - * Marvel

Input what you're in the mood for here:

MovieRecs

Your Top Recommended Movies Are:

title	plot	genre	score	tomatometer_rating	audience_rating	imdb_score
0 Florence Foster Jenkins	In the 1940s, New York socialite Florence Foster Jenkins (Meryl Streep) dreams of becoming a great opera singer. Unfortunately, her ambition far exceeds her talent. The voice Florence hears in her head is beautiful, but to everyone else it is quite lousy. Her husband St. Clair goes to extreme lengths to make sure his wife never finds out how awful she truly is. When Florence announces her plans for a concert at Carnegie Hall, St. Clair soon realizes that he's facing his greatest challenge yet.	Comedy, Drama, Musical & Performing Arts	73.71	87.0	68.0	68.0
1 The Devil Wears Prada	Andy (Anne Hathaway) is a recent college graduate with big dreams. Upon landing a job at prestigious Runway magazine, she finds herself the assistant to diabolical editor Miranda Priestly (Meryl Streep). Andy questions her ability to survive her grim tour as Miranda's whipping girl without getting scorched.	Comedy, Drama	72.77	75.0	76.0	69.0
2 Julie & Julia	Frustrated with a soul-killing job, New Yorker Julie Powell (Amy Adams) embarks on a daring project: she vows to prepare all 524 recipes in Julia Childs' landmark cookbook, "Mastering the Art of French Cooking." Intertwined with Julie's story is the true tale of how Julia Child (Meryl Streep) herself conquered French cuisine with passion, fearlessness, and plenty of butter.	Comedy, Drama	71.69	77.0	70.0	70.0
3 Hope Springs	Long-married couple Kay (Meryl Streep) and Arnold (Tommy Lee Jones) love each other, but after so many years together, Kay feels the need to spice things up and reconnect with her husband. Hearing of a renowned marriage counselor (Steve Carell) in the town of Great Hope Springs, Kay manages to persuade her stubborn mate to attend a retreat with her. As they shed their bedroom hang-ups, Kay and Arnold find that the real challenge is reigniting the spark that originally brought them together.	Comedy, Drama	63.95	75.0	55.0	63.0
4 Mamma Mia!	Donna (Meryl Streep), an independent hotelier in the Greek Islands, is preparing for her daughter's wedding with the help of two old friends. Meanwhile Sophie, the spirited bride, has a plan. She secretly invites three men from her mother's past in hope of meeting her real father and having him escort her down the aisle on her big day.	Comedy, Musical & Performing Arts, Romance	60.85	55.0	66.0	64.0
5 It's Complicated	Jane (Meryl Streep), a successful restaurateur, has been divorced from Jake (Alec Baldwin) for many years, although they remain friends. When an innocent meal together turns into an affair, Jane finds that she is the other woman in the now-remarried Jake's life. Meanwhile, Adam (Steve Martin), an architect hired to remodel her kitchen, is still stinging from his own divorce. He finds himself falling in love with Jane but soon realizes that he has become part of a triangle.	Comedy, Romance	60.80	58.0	60.0	65.0

THANKS!

Jen Zieger

jenczieger@gmail.com

CREDITS: This presentation template was created by [Slidesgo](#), including icons from [Flaticon](#), infographics & images by [Freepik](#)

Please keep this slide as attribution



RESOURCES

Images:

- IMDB Logo: <https://en.wikipedia.org/wiki/IMDb>
- Rotten Tomatoes Logo:
https://commons.wikimedia.org/wiki/File:Rotten_Tomatoes_logo.svg
- <https://www.freepik.com/vectors/man> Man vector created by macrovector - www.freepik.com
- <https://www.freepik.com/vectors/people> People vector created by freepik - www.freepik.com
- <https://www.freepik.com/vectors/vintage> Vintage vector created by rawpixel.com - www.freepik.com
- <https://www.freepik.com/vectors/background> Background vector created by macrovector - www.freepik.com
- <https://www.freepik.com/vectors/people> People vector created by freepik - www.freepik.com
- <https://www.freepik.com/vectors/people> People vector created by pch.vector - www.freepik.com
- <https://www.freepik.com/photos/house> House photo created by freepik - www.freepik.com

Datasets:

- Rotten Tomatoes Dataset:
<https://www.kaggle.com/stefanoleone992/rotten-tomatoes-movies-and-critic-reviews-dataset>
- IMDb Dataset:
<https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset?select=IMDb+movies.csv>

Supplementary Information:

- Similarweb
- 78% Link
- Check images and remove unused.