

AMAZON MOVIE REVIEWS RECOMMENDATION SYSTEM

Team 3

Schinria Islam

Yuchen Shi

Yanran Wang

Yi Zhou

RECOMMENDATION SYSTEM

- PART I
 - User-based Recommendation
- PART II
 - Product-based Recommendation (GENRE DETECTION)

NEXTFOLK

OUR RECOMMENDATION SYSTEM

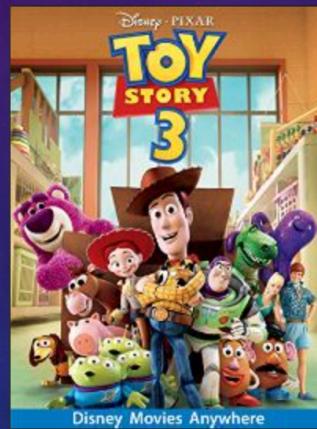
- PART I
 - User-based Recommendation
- PART II
 - Product-based Recommendation (GENRE DETECTION)

NEX~~T~~FLIK

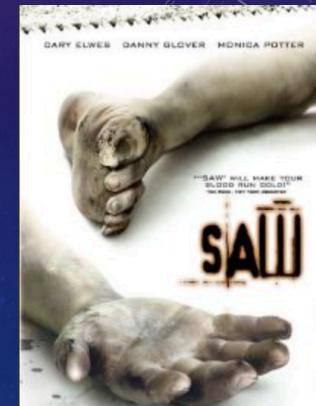
DR. ZHENG'S MOVIE PREFERENCES

NEXTFLIK

- Iron Man
- Science Fiction, Adventure, Action
- Dr. Zheng Rating: ★★★



- True Grit
- Drama, Western, Adventure
- Dr. Zheng Rating: ★★★★★



- Batman
- Drama, Action
- Dr. Zheng Rating: ★★

- Toy Story 3
- Fantasy, Adventure, Comedy, Kids & Family
- Dr. Zheng Rating: ★★★★

- Saw
- Mystery, Horror
- Dr. Zheng Rating: ★★★

THE GOAL: PART I



- Step 1: Find the users have similar taste based on input
- Step 2: Use similar users' movie list to find the best movie to recommend



HOW TO FIND SIMILAR USERS

- Make out users who have watched at least two of the five input movie.
- Delete strange users: who commented on more than 3 movies a week.
- Calculate the $1/(Euclidean\ Distance)$ to describe how similar they are to the input.
- Output: Similar users' ID and their similarity score

SCORING MATRIX



Get all the movies similar users have written reviews and delete those have an illegal product ID. Let's say there are n movies in total...

Part A: Options based on adjusted scores

- Scale the score for reviews of all these movies:
- adjusted score of the review = $\text{score} * \exp(\text{helpfulness}) * \log(\text{denominator of helpfulness} + 2)$

1. Get the mean of adjusted score of the review for all movies
2. Rank all n movies based on adjusted score and get the first 200

Part B: Options based on number of comments

1. Rank all n movies based on the number of comments

Part C: kind of combining A and B

1. Get the first 10% of movies based on B's rank
2. Rank the first 10% movies based on adjusted score and get the first 200

SCORING MATRIX



Then combine A,B and C to get 600 movies...

1. Delete duplicate and bad movies
 - What we mean by “bad” movies
 2. Create a score matrix for movies left, row is similar viewer ID, column is product ID for movies
 3. Time each similar viewer’s score to the score matrix
-
- **How do we make recommendations using this method?**
 - Calculate the mean score for each movie, recommend five movies with highest score

SCORING MATRIX

	movie 1	movie 2	movie 3	movie 4
User 1	4	3	NA	4
User 2	NA	4	5	3

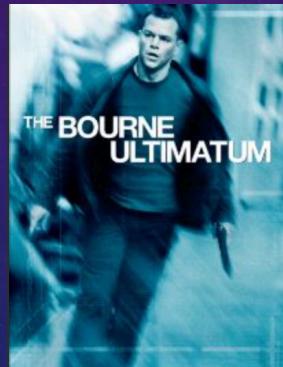
×

	Similarity	Similarity	Similarity	Similarity
User 1	0.5	0.5	0.5	0.5
User 2	0.7	0.7	0.7	3

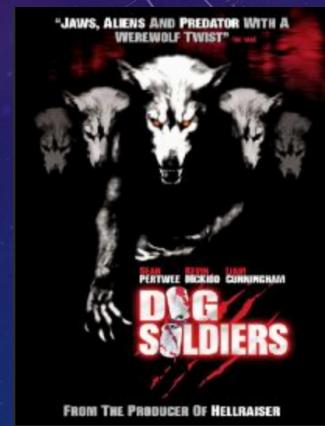
- How do we make recommendations using this method?
 - Calculate the mean score for each movie, recommend five movies with highest score

DR. ZHENG, NEXT YOU SHOULD WATCH: *Recommendation System 1*

- Grindhouse: Planet Terror
- Action, Comedy, Horror
- Amazon Rating: ★★★★



- Team America
- Action, Comedy
- Amazon Rating: ★★★★



- Mission Impossible III
- Thriller, Adventure, Action
- Amazon Rating: ★★★★

- The Bourne Ultimatum
- Thriller, Action
- Amazon Rating: ★★★★

- Dog Soldiers
- Fantasy, Action, Horror
- Amazon Rating: ★★★★

OUR RECOMMENDATION SYSTEM

- PART I
 - User-based Recommendation
- PART II
 - Product-based Recommendation (GENRE DETECTION)

NEX~~T~~FLIK

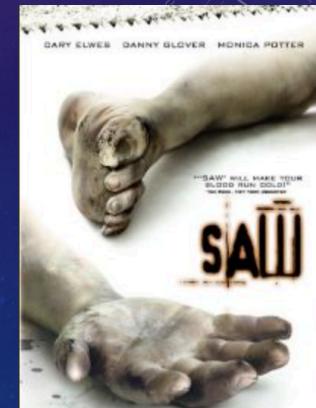
11 DR. ZHENG'S MOVIE PREFERENCES

NEXTFLIK

- Iron Man
- Science Fiction, Adventure, Action
- Dr. Zheng Rating: ★★★



- True Grit
- Drama, Western, Adventure
- Dr. Zheng Rating: ★★★★★



- Batman
- Drama, Action
- Dr. Zheng Rating: ★★

- Toy Story 3
- Fantasy, Adventure, Comedy, Kids & Family
- Dr. Zheng Rating: ★★★★

- Saw
- Mystery, Horror
- Dr. Zheng Rating: ★★★

27 AMAZON MOVIE GENRES...



Indie & Art House



Music Videos & Concerts



Science Fiction



Westerns



African American Cine-



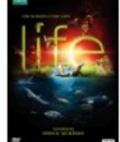
Boxed Sets



Documentary



Exercise & Fitness



Foreign Language & In-



Animation



Classics



Drama



Faith & Spirituality



Gay & Lesbian



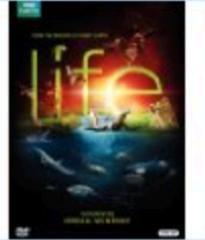
Holiday & Seasonal



Kids & Family



Mystery & Thrillers



Special Interests

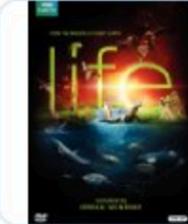
Action & Adventure



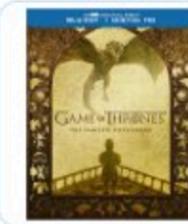
Anime



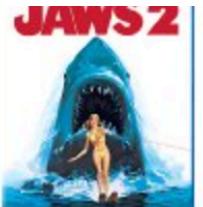
Comedy



Educational



Fantasy



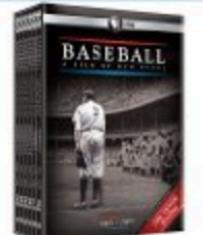
Horror



Military & War



Romance



Sports

...27 GENRE LEXICON DICTIONARIES

MUSIC
VIDEOS &
CONCERTS
music
video
concert
danc
production
stage
broadway
theater
song
show
choreo
sing
orchestra
sound
vocal
costume
number
rock
disco
classical
cappella
instrument
never-before
discography
season
complete
series
popular
commemorat
collect
pack

EXERCISE & FITNESS

exercis
fitness
fit
muscle
tone
shape
thin
abs
rock-hard
physique
weight
lean
train
hard
workout
pump
inspir
body
build
gym
rat
at-home
iron
dance
sequence
cardio
aerobic

BOXED SETS

box
set
compilation
books
recording
film
package
anthology
essential
never-before
discography
season
complete
series
popular
commemorat
collect
pack

WESTERNS

west
american
old
cowboy
revolve
rifle
horse
boot
native
bounty
hunter
outlaw
settler
town
spur
cavalry
ranch
sling
saloon

COMEDY

comedy
funny
comic
laugh
satir
humor
sketch
fancy
excite
entertain
parody
mock
lightheart
pun
amuse
stand-up
satire
screwball
joke
improvise
farce
sitcom
camp
fun
hilar
merry

MYSTERY & THRILLER

myster
thriller
detect
evil
puzzle
police
cruel
suspens
secre
riddle
crypt
perplex
tease
whodunit
crim
twist
psycho
fascinat
mind
trap
accident
investig
thief
disappear
intense
disturb
horify

FANTASY

fantasy
magic
myth
supernatural
folk
exotic
world
wonder
escap
extraordinary
sword
sorcer
ring
nether
fairy-tale
legend
imagin
dream
hallucinat
mystic
power
superhuman
force
fly
dragon
bizarre
whim
accident
investig
thief
disappear
intense
disturb
horify

HORROR

horror
kill
scream
monster
torture
terror
demon
shark
unsettl
dread
fright
panic
alarm
fear
terrify
shock
captiv
macabre
supernatural
unknown
ghost
extraterrestrial
vampire
werewolve
gore
vicious
evil
witch
zombie

ACTION & ADVENTURE

action
adventur
risk
explosi
fight
daring
dare
escap
hero
blood
brother
duty
honor
violence
warrior
spy
mission
gun
gadget
danger
stunt
alert
plan
dash
bustle
power
treasure
map
fear
island
journey
experience
feat
venture
deed

THE GOAL: PART II



- Step 1: Remove movies that didn't get matched with any genre by text and summary content
 - 7,911,684 reviews → 7,887,702 reviews
- Step 2: Calculate mean genre detection scores per movie and join genre detection data with product ID table

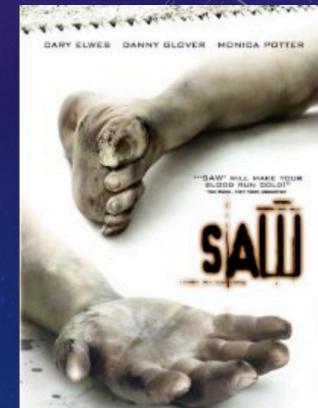
DR. ZHENG'S MOVIE PREFERENCES



- Iron Man
- Science Fiction, Adventure, Action
- Dr. Zheng Rating: ★★★



- True Grit
- Drama, Western, Adventure
- Dr. Zheng Rating: ★★★★★



- Batman
- Drama, Action
- Dr. Zheng Rating: ★★

- Toy Story 3
- Fantasy, Adventure, Comedy, Kids & Family
- Dr. Zheng Rating: ★★★★

- Saw
- Mystery, Horror
- Dr. Zheng Rating: ★★★

16

DR. ZHENG'S MOVIE PREFERENCES

- Iron Man
- Science Fiction, Adventure, Action
- Dr. Zheng Rating: ★★★



- True Grit
- Drama, Western, Adventure
- Dr. Zheng Rating: ★★★★★



NEXTFLIK



- Batman
- Drama, Action
- Dr. Zheng Rating: ★★

- Toy Story 3
- Fantasy, Adventure, Comedy, Kids & Family
- Dr. Zheng Rating: ★★★★

- Saw
- Mystery, Horror
- Dr. Zheng Rating: ★★★

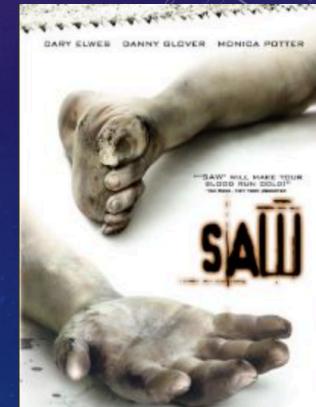
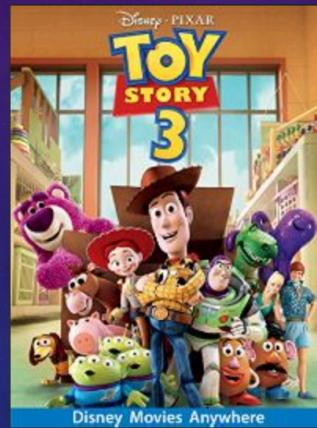
COMPARING GENRES



- Step 3: Extract actual movie genres for user's inputted movies using Amazon API
- Step 4: Calculate genre-proportions for each product ID for Amazon API genres
- Step 5: Similarity scores

18

DR. ZHENG, LIKE THESE MOVIES?



- Batman
- Drama, Action
- Dr. Zheng Rating: ★★

- Toy Story 3
- Fantasy, Adventure, Comedy, Kids & Family
- Dr. Zheng Rating: ★★★★

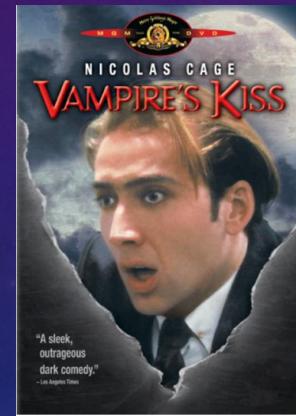
- Saw
- Mystery, Horror
- Dr. Zheng Rating: ★★★

NEXTFLIK

DR. ZHENG, YOU SHOULD WATCH:

Recommendation System 2

- The Ghost Breakers
- Horror, Mystery & Thrillers, Comedy
- Amazon Rating: 



- The Good Witch
- Kids & Family
- Amazon Rating: 



- The Monster That Challenged the World
- Science Fiction, Horror
- Amazon Rating: 

- Vampire Kill
- Fantasy, Horror, Mystery & Thrillers
- Amazon Rating: 

- Dawn of the Living Dead
- Science Fiction, Horror
- Amazon Rating: 

BAD MOVIES

- Type 1
 - Rating histogram follow a shape of “C”, which means there are as many 5 rating as 1 rating and they are both near 30%~40% over the total sum of the reviews
 - Hidden transactions between customers and movie sellers
 - Lots of 5's and 1's not reasonable
 - Low ratings basically about the movie, not product quality or sale management issues
 - Outliers
 - Comparison with Rotten Tomatoes and the famous Chinese website Douban

DIFFERENT RATING STRATEGIES BETWEEN CHINESE AND AMERICAN CUSTOMERS

Search Results - Rotten Amazon.com: Customer ÁjÁjÁ ÁjÁjÁ Á... 13:37

www.amazon.com/Tree-Three-Disc-Blu-ray-Combo-Digital/product-reviews/B005HV6Y5W/ref=cm_cr_dp_see_all_summary?ie=UTF8 读

amazon Try Prime sp 50% 0.1K/s + 0.1K/s

All Hello, Sign in Your Account Try Prime Lists Cart 0

Departments Your Amazon.com Today's Deals Gift Cards Sell Help

The Tree of Life (Three-Disc Blu-ray/DVD Combo + Digital Copy) > Customer Reviews

Customer Reviews 858

3.3 out of 5 stars 45% 10%

5 star 45% 4 star

4 star 10% 3 star

3 star 8% 2 star

2 star 8% 1 star

29% **Rate this item** **Write a review**

The Tree of Life (Three-Disc Blu-ray/DVD Combo + Digital Copy) by Brad Pitt

Format: Blu-ray | Change
Price: \$8.91 + Free shipping with Amazon Prime

Add to Cart
 Add to Wish List

RATINGS ON DOUBAN WEBSITE

Screenshot of a web browser showing the Douban movie rating page for "The Tree of Life (2011)".

The page includes the following information:

- Movie Title:** 生命之树 The Tree of Life (2011)
- Director:** 泰伦斯·马可克
- Cast:** 布拉德·皮特 / 西恩·潘 / 杰西卡·查斯坦 / 卡瑞·玛切特 / 费奥纳·肖 / 更多...
- Type:** 剧情 / 家庭 / 奇幻
- Official Website:** www.twowaysthroughlife.com
- Production Country:** 美国
- Language:** 英语
- Release Dates:** 2011-05-16 (戛纳电影节) / 2011-05-27 (美国部分上映)
- Runtime:** 139分钟
- Aliases:** 生命树(港) / 永生树(台) / 家谱 / Tree of Life
- IMDb Link:** tt0478304

Douban Rating: 6.9 (27299人评价)

Rating	Percentage
5星	16.1%
4星	30.9%
3星	38.7%
2星	10.7%
1星	3.7%

Comparison: 好于 27% 剧情片
好于 31% 家庭片

Where to Watch:

- 腾讯视频: 付费 5 元
- 优酷视频: 付费

Advertisement: 垫底辣妹 Flying Colors 4月14日全国公映 世界上最大的谎言是你不行

QR Code: Douban logo with the slogan: 和有趣的人做有趣的事

Bottom Navigation: Windows icons, browser tabs (Safari, Firefox, Microsoft Edge, Google Chrome, Internet Explorer), and system status bar (GTGRAPHICS.DE, ENG, 15:44, 2016/4/11).

RATING OF FRESHNESS ON ROTTEN TOMATOES

Search Results - Rotten a Amazon.com: Customer

www.rottentomatoes.com/search/?search=The+Tree+of+Life

PLAY FOR FREE

MOVIES

Score	Movie Title	Year	Director	Cast
84%	The Tree of Life	(2011)		Brad Pitt, Sean Penn, Jessica Chastain
48%	The Tree of Life	(2008)		Alessandra Veronesi, Elio Toaff, Hava Volterra
No Score Yet	Charles Darwin and the Tree of Life	(2009)		David Attenborough, David Attenborough, Jeremy Bristow
No Score Yet	The Life Of Trees	(2012)		
No Score Yet	The Olive and the Tree: The Secret Life of the Druze			

WARS

MUSIC BY JOHN WILLIAMS

Buy on iTunes

https://www.googleadservices.com/pagead/aclk?sa=L&ai=C7P52no4KV4L4I8SPhASw0LqIBO_ox6hEg6DPvsgCpOrH5tkEEAEgn8arIWDJxqmLwKTYD6AB_evzwwPIAQLgAgCoAwHIA5kEqgS4AU_QkGsybRAu3j8ywohumU5Rs3CmOJB44...

GTGRAPHICS.DE 13:42 2016/4/10

COMPARISON

- Ratings from Chinese viewers follow a shape of “D” instead for this move.
 - Imagine that if you see the pattern on the website, you are probably going to be a little interested in the movie since there are actually not too few people like it.
- Also rating on rottentomatoes.com is surprisingly high!

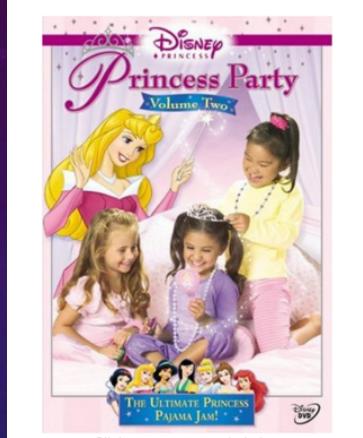
CONCLUSION

1. Amazon customers tend to be mean „ಠ_ಠ“
2. Chinese viewers tend to give 4 instead of 5 if they like it, and give 3 or 2 instead of 1 when they don't like it. May be they are too shy to expose what they are actually thinking about too much...눈_눈

BAD MOVIES

- Type 2
 - Rating histogram follow a shape of “b”, which means there are as many 3 rating as 1 rating and they are both near 30%~40% over the total sum of the reviews.
 - Movies maybe aren't' terrible but also not eye-catching

LOOK AT THIS MOVIE ↵_↵



CONCLUSIONS

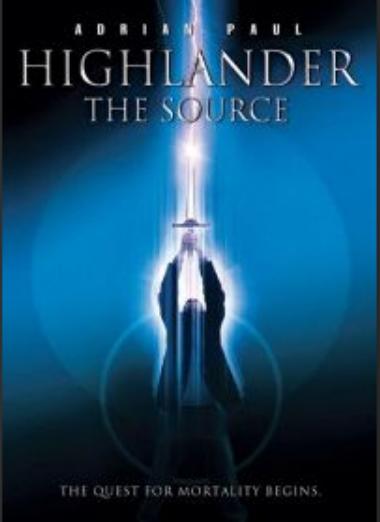
Maybe you think this movie “Disney Princess Party” is not too bad since the average score is 2.7. However most of the people are giving 1 and 3 for rating, which may suggest that this movie is not that good.

BAD MOVIES

- Type 3
 - Rating histogram follow a shape of “L”, which most of the ratings are 1.
 - This pattern may suggest that this movie is as bad as you can imagine...

FOR EXAMPLE..... OUCH.

Highlander: The Source 2007 R cc



The poster features a dark figure standing in a bright, glowing blue light, with the title "HIGHLANDER THE SOURCE" and the tagline "THE QUEST FOR MORTALITY BEGINS."

★★★★★ (164) | IMDb 3.1/10

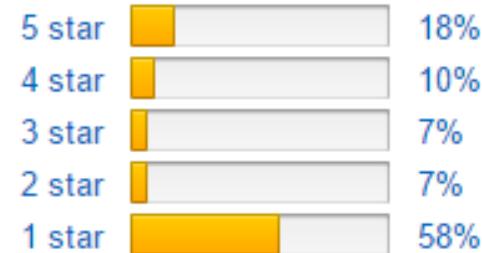
The world is falling into chaos and Duncan MacLeod before the love of his life left.

Starring: Adrian Paul, Thekla Reuten
Runtime: 1 hour, 26 minutes
Available to watch on [supported devices](#).

Customer Reviews

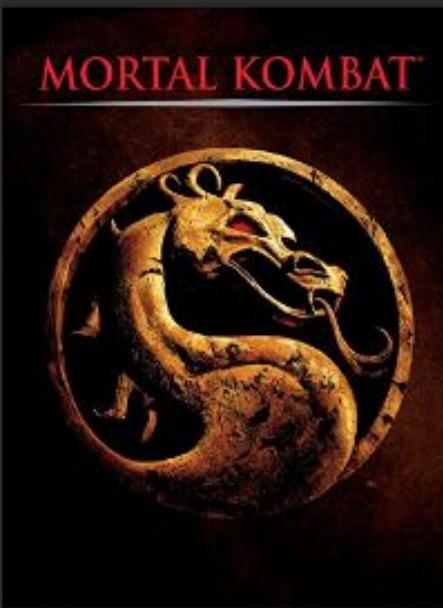
★★★★★ (164)

2.2 out of 5 stars ▾



FINALLY...SOMETHING STRANGE

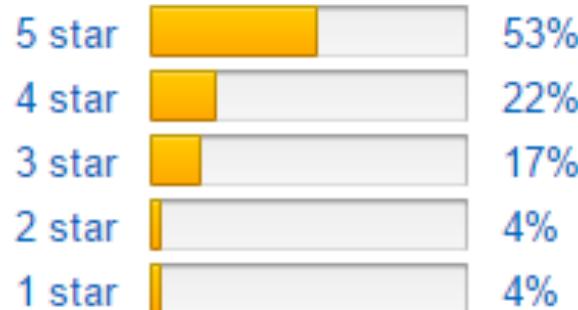
Mortal Kombat 1995



Customer Reviews

★★★★☆ 335

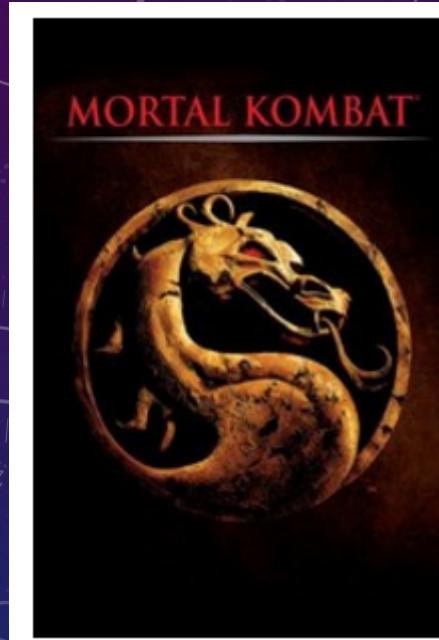
4.2 out of 5 stars ▾



THE CLASSICAL TYPE OF “GOOD” MOVIE

- High rating with many 5, 4, 3, looking like a shape of “F”
- It's not difficult to infer that people give a lot of 5, 4 and 3 rating, which seems pretty “good”.

HOWEVER...



TOMATOMETER

All Critics | Top Critics



33%

Average Rating: 4.5/10
Reviews Counted: 30
Fresh: 10
Rotten: 20

Critics Consensus: Despite an effective otherworldly atmosphere and appropriately cheesy visuals, Mortal Kombat suffers from its poorly constructed plot, laughable dialogue, and subpar acting.

AUDIENCE SCORE



58%

liked it

Average Rating: 2.9/5
User Ratings: 311,133

ADD YOUR RATING



+ WANT TO SEE

NOT INTERESTED



Add a Review (Optional)

TAKEAWAYS: BLOOD AND TEARS

- Don't trust ratings on Amazon totally. They may cheat on you.
- For this movie it's not easy for us to get rid of, however, since the rating pattern looks great.
- As a result, why not just keep this movie. Some people just love it!!!

TAKEAWAYS: BLOOD AND TEARS

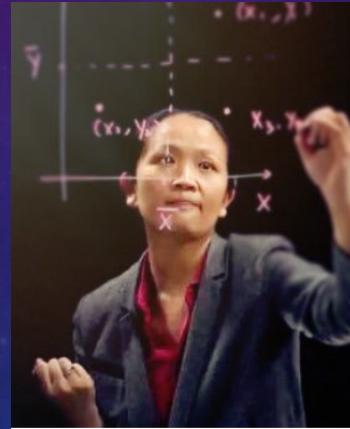
- Don't trust ratings on Amazon totally. They may cheat on you.
- For this movie it's not easy for us to get rid of, however, since the rating pattern looks great.
- As a result, why not just keep this movie. Some people just love it!!!



- Most the movies falling in the top 3 bad movie types are rated “R”

Thank you!

Our team's personal favorite movie:



- Data Science with Dr. Zheng: Made Easy!
- Action, Adventure, Comedy, Educational
- Team 3's Rating: ★★★★★