Review Sentiment Analysis

Matthew Lucht

Overview

The goal of this project was to build a machine learning review analysis pipeline that analyzes the sentiment of movie reviews and identifies key themes in positive and negative reviews.

To accomplish this, these elements were used:

1. SVM classifier
2. TF-IDF
3. GloVe
4. NMF
5. OpenAI API

Real World Applications

- Consumer tool to get quickly informed on the general population's opinion on media without having to read several lengthy reviews

- Also provides the benefit of easy access to the most frequent praises and complaints of the movie

- Producer tool to get quickly informed on customer and market feedback, providing an easy way to see exactly what people do and don't like about the movie

- General idea of this pipeline can be easily expanded to things like amazon product reviews or yelp store reviews, letting the makers / owners know what they need to fix to improve customer satisfaction

Pipeline part 1.

1. Big review datasets from kaggle

↓

2. Data preprocessing

↓

3. Input: Clean large dataset file

↓

4. SVM classifier training

+

5. GloVE sentiment analysis fitting

↓

6. Outputs: Model to classify and analyze review set

Pipeline part 2.

1. Specific movie reviews scraped from web
2. Data formatting and preprocessing
3. Input: Clean small dataset file
4. SVM classification into positive and negative
5. GloVe sub-categorization
6. Analysis of common review topics / themes with NMF
7. Outputs: OpenAI generated summary

Accuracy of SVM and GloVe

- SVM is used for the initial split of reviews, with 90% accuracy, and then GloVe is used torm the strong and weak subcategories with a lower confidence level
- This ensures minimal mistakes in classifying a positive as negative, which was happening much more when I started with 4-way classification

	precision	recall	f1-score	support
negative	0.90	0.90	0.90	5000
positive	0.90	0.90	0.90	5000
accuracy			0.90	10000
macro avg	0.90	0.90	0.90	10000
weighted avg	0.90	0.90	0.90	10000

=== Strong vs. Weak Negative ===
Accuracy: 0.7254

	precision	recall	f1-score	support
strong negative	0.62	0.09	0.15	618
weak negative	0.73	0.98	0.84	1560
accuracy			0.73	2178
macro avg	0.67	0.53	0.49	2178
weighted avg	0.70	0.73	0.64	2178

=== Strong vs. Weak Positive ===
Accuracy: 0.6422

	precision	recall	f1-score	support
weak positive	0.61	0.31	0.41	1791
strong positive	0.65	0.87	0.74	2642
accuracy			0.64	4433
macro avg	0.63	0.59	0.58	4433
weighted avg	0.63	0.64	0.61	4433

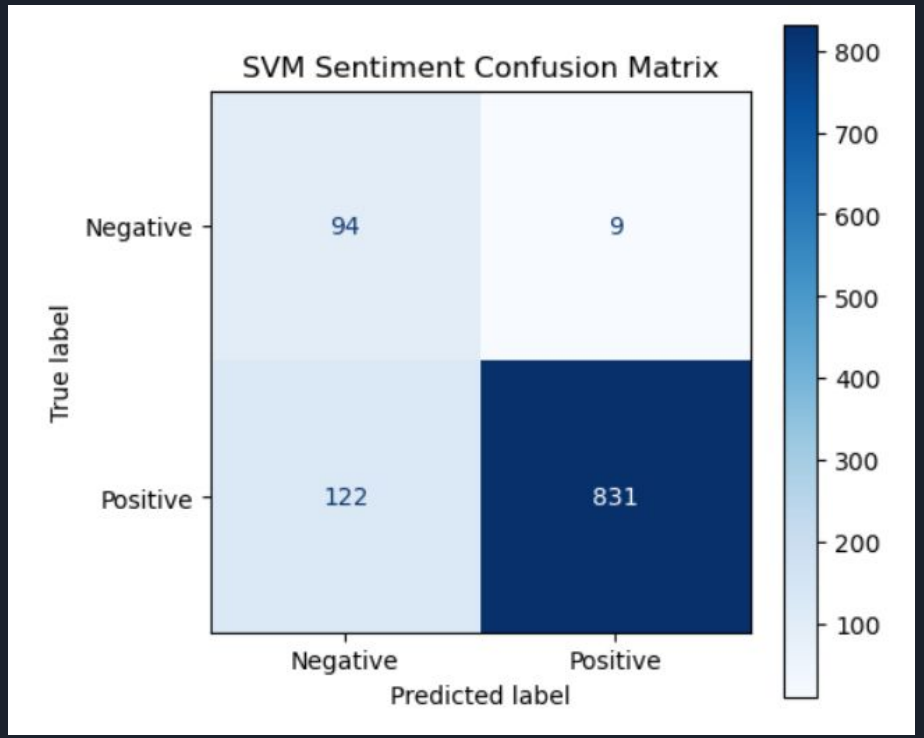

JASON MOMOA
JACK BLACK
DANIELLE BROOKS
EMMA MYERS
SEBASTIAN HANSEN
A
MINECRAFT
MOVIE
ONLY IN THEATERS
APRIL 4
IN Dolby Cinema REAL D 3D AND IMAX
PG
LEGENDARY


TIMOTHÉE CHALAMET
ZENDAYA
REBECCA FERGUSON
JOSH BROLIN
AUSTIN BUTLER
FLORENCE PUGH
DAVE BAUTISTA
CHRISTOPHER WALKEN
LÉA SEYDOUX
SOUHEILA YACOUB
STELLAN SKARSGÅRD
CHARLOTTE RAMPLING
JAVIER BARDEM
DUNE
PART TWO
IN CINEMAS
MARCH 1
LEGENDARY
FILMED FOR IMAX

Testing on Dune part 2

- The initial SVM classification was highly accurate, but the follow up sub-categorization misses some strong negative.
- This could be due to the training data and review data lacking strong negative reviews, making them the least common review type

	precision	recall	f1-score	support
0	0.44	0.91	0.59	103
1	0.99	0.87	0.93	953
accuracy			0.88	1056
macro avg	0.71	0.89	0.76	1056
weighted avg	0.94	0.88	0.89	1056



Testing on Dune part 2

```
Strong negative:  2
Weak negative:  214
Weak positive:  49
Strong positive: 791
```

- Words that show up frequently in multiple categories of reviews are filtered out, as they are likely not informative of sentiment

- The words identified in the positive and negative categories are good to see, as they match the positive and negative sentiment and pass the subjective test of what people mentioned about the movie in the reviews I read

```
Common to ALL categories (21): sci fi, felt, acting, austin butler, book, baron, austin, action, sci, denis,
```

```
Distinctive in weak negative (70): slow, boring, make, sequel, hour, pace, look, long, alia, changes,
```

```
Distinctive in strong positive (72): better, amazing, slow, end, loved, narrative, cinematic, depth,
```

Testing on Dune part 2

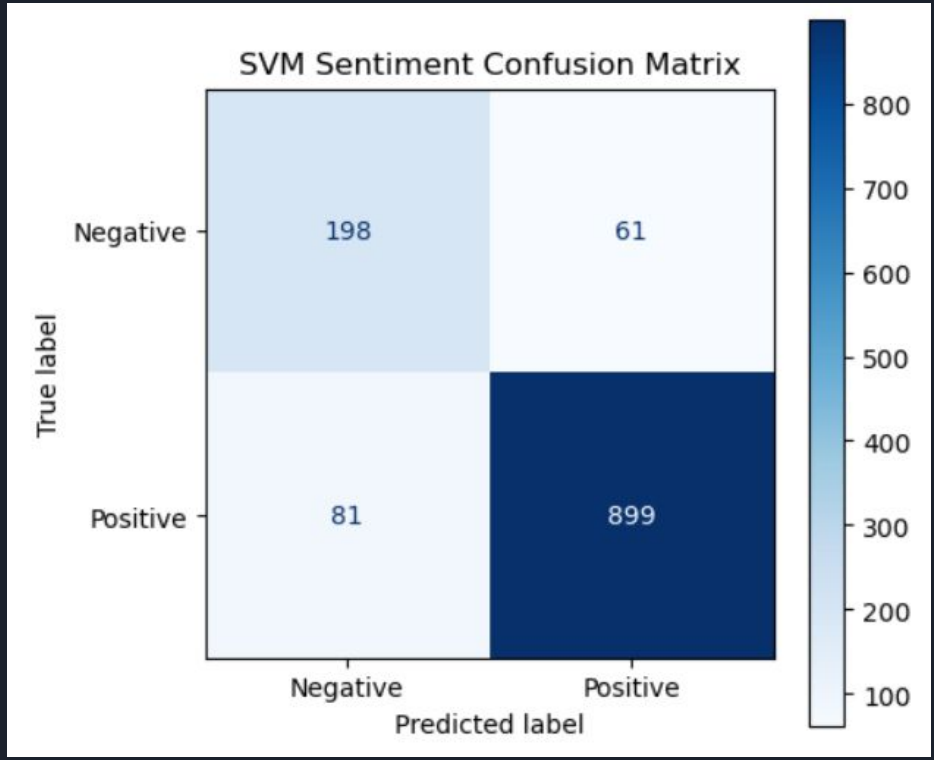
- The score calculated by weighting strong and weak negative and positive reviews is highly accurate to the actual data, and a good representation of the score on the website.

- This is a big improvement from calculating the score using only the number of binary negative vs. positive reviews

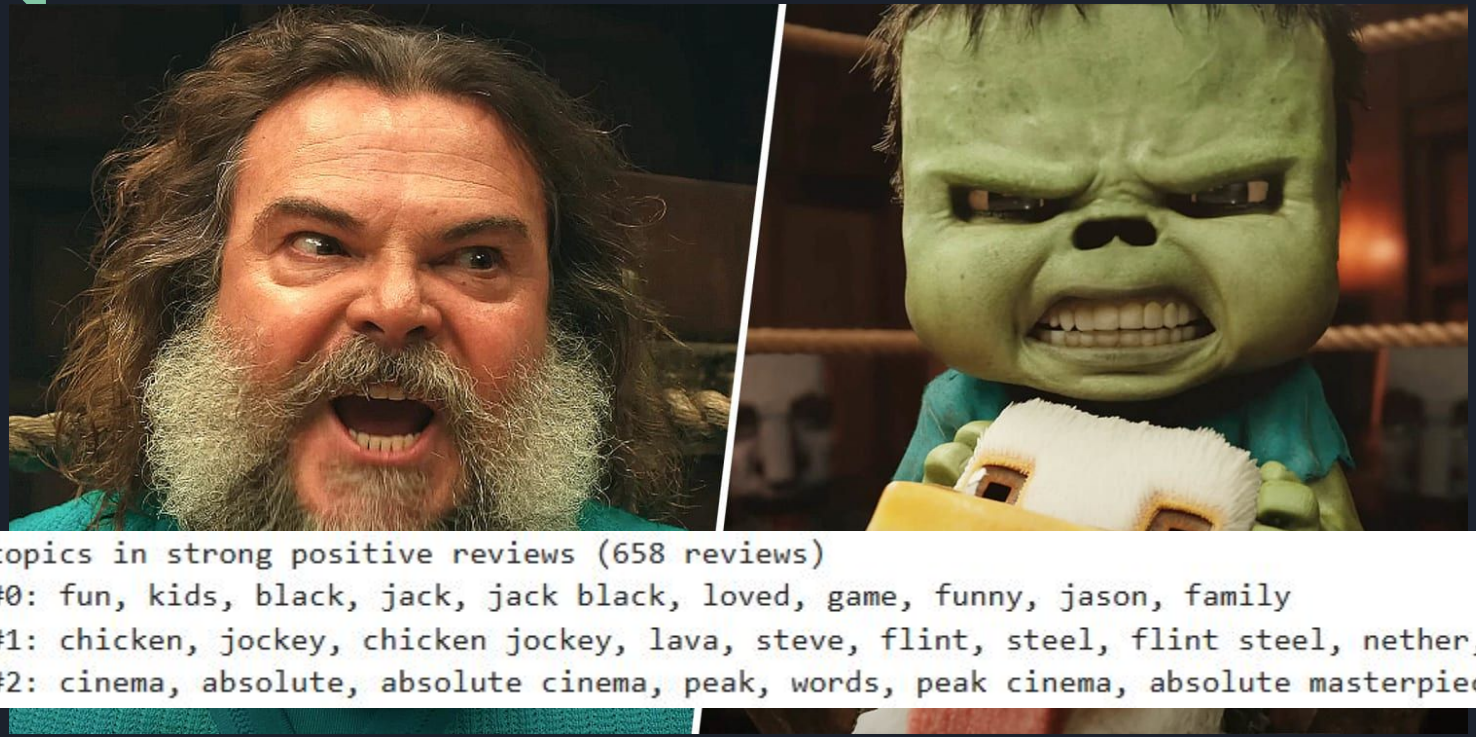
```
Total reviews: 1056
Breakdown by subsentiment:
  strong positive: 791
  weak negative  : 214
  weak positive  : 49
  strong negative: 2
Predicted average review score: 8.47
Actual average review score: 8.58
```

Testing on Minecraft Movie

	precision	recall	f1-score	support
0	0.71	0.76	0.74	259
1	0.94	0.92	0.93	980
accuracy			0.89	1239
macro avg	0.82	0.84	0.83	1239
weighted avg	0.89	0.89	0.89	1239



Testing on Minecraft Movie



```
Top 5 topics in strong positive reviews (658 reviews)
Topic #0: fun, kids, black, jack, jack black, loved, game, funny, jason, family
Topic #1: chicken, jockey, chicken jockey, lava, steve, flint, steel, flint steel, nether, lava chicken
Topic #2: cinema, absolute, absolute cinema, peak, words, peak cinema, absolute masterpiece, masterpiece,
```

Testing on Minecraft Movie

- In this example, the review score calculated was even closer to the data, but it was far off the score on the website

- This shows a disconnect between the review scores of people writing reviews and those just leaving star ratings, which could be a problem with the model to solve in the future.

```
Total reviews: 1239
  strong positive: 658
  weak positive  : 302
  weak negative  : 260
  strong negative: 19
Predicted average review score: 7.61
Actual average review score: 7.63
```

Generating Review Summaries

- Using the OpenAI API, the ipynb notebook is able to send a request to OpenAI's server to get a response from ChatGPT with the given prompt

- The summary is formatted to be quick to read with all the most relevant information easily available, and should incorporate the most prominent themes of both positive and negative themes of reviews

```python
# Make chatGPT review summary using top topics from each sentiment category
def generate_overall_summary(
    strong_pos_topics, weak_pos_topics,
    weak_neg_topics,   strong_neg_topics
):
    prompt = f"""
You are a movie review summarizer.  Write one concise paragraph (2-3 sentences) that:
1. Starts with the strongest positives (themes: {', '.join(strong_pos_topics)}).
2. Then covers the weaker positives (themes: {', '.join(weak_pos_topics)}).
3. Follows with the weaker negatives (themes: {', '.join(weak_neg_topics)}).
4. Ends on the strongest negatives (themes: {', '.join(strong_neg_topics)}).
Keep it punchy and flow naturally from positive to negative.
"""
    resp = openai.chat.completions.create(
        model="gpt-4o",
        messages=[
            {"role": "system", "content": "You summarize movie reviews."},
            {"role": "user",   "content": prompt}
        ],
        max_tokens=120,
        temperature=0.7,
    )
    return resp.choices[0].message.content.strip()
```

Movie Review Highlights

"Minecraft: The Movie" is an absolute masterpiece that delivers peak cinema for both kids and adults, capturing the essence of the beloved video game with its blocky charm and creative adventure. With standout performances from Jack Black and Jason Momoa, the film offers a funny and enjoyable ride that fans of the game will definitely appreciate, while the impressive CGI and humor make it a delightful experience for families. However, despite its fun and familiar elements, the plot can feel a bit formulaic and some scenes fall short of the expected quality, leaving a few fans feeling it wasn't the cinematic masterpiece they hoped for.

Movie Review Highlights

Denis Villeneuve's adaptation of "Dune" is a stunning visual and cinematic masterpiece that brings Frank Herbert's intricate sci-fi world to life with remarkable depth and authenticity. The performances by Timothée Chalamet, Zendaya, and Rebecca Ferguson are compelling, and the film's sound design and Hans Zimmer's score enhance the epic narrative, making it a must-see on the big screen. However, some may find the pacing slow and the adaptation of the complex source material a bit challenging to follow, with certain plot elements feeling rushed or underexplained, potentially leaving those unfamiliar with the book lost.

Testing my Pipeline vs. LLM

- I input a “hidden” csv file, where the score column was deleted leaving only the reviews and review titles

- Asked ChatGPT: “Without using the internet, only reading from this data, formulate an estimate for the average review score of this film, and give a brief summary of the most common positive and negative themes.”

- Tested for both films, with ChatGPT 4o (basic) and 4.5 (advanced) models

ChatGPT 4o

Top Positive And Negative Words

		word	frequency	sentiment
1	606	movie	605	positive
2	239	dune	427	positive
3	330	film	343	positive
4	80	best	251	positive
5	847	story	206	positive
6	394	great	191	positive

Estimated average review score: ~6.5/10

Summary of themes:

- **Most common positives:** Viewers praised the *movie*, its *story*, and described it as the *best film*, often mentioning *Dune* specifically. This suggests strong appreciation for the storytelling, quality of production, and its status as a standout science fiction film.
- **Most common negatives:** Common negative themes involved mentions of the same core elements, such as *movie* and *story*, indicating that some viewers had mixed or contrasting opinions on the film's pacing, clarity, or adaptation quality.

ChatGPT 4.5

Estimated Average Review Score:

Approximately 6.6 out of 10

Common Positive Themes:

- Reviewers frequently praised the film as a "**masterpiece,**" noting it as a "**phenomenal**" experience with strong emphasis on impressive **visuals** and **storytelling**.
- Terms like "**epic**" and "**spectacle**" highlight appreciation for grand scale and visually stunning cinematography.

Common Negative Themes:

- Criticism often focused on the film feeling "**slow**" and "**long,**" suggesting pacing issues were a significant downside.
- Some described the film as "**soulless,**" "**boring,**" or criticized the lack of engagement and insufficient excitement, indicating that despite visual brilliance, emotional depth and narrative pacing were common complaints.

Dune Part 2

ChatGPT 4o

Top Minecraft Review Words

		word	frequency	sentiment
1	612	movie	498	positive
2	597	minecraft	157	positive
3	399	good	124	positive
4	404	great	114	positive
5	531	like	96	positive
6	496	just	88	positive

Estimated average review score: ~6.2/10

Summary of common themes:

- **Positive themes:** Reviewers frequently praised the *movie* as *good* or *great*, often referring positively to *Minecraft* and using words like *fun*, *family*, and *entertaining*. This suggests audiences found the film enjoyable, especially for fans or family settings.
- **Negative themes:** Some negative feedback centered around the *movie* being *boring*, *bad*, or a *waste*, indicating disappointment with execution, pacing, or expectations from the source material.

ChatGPT 4.5

Estimated Average Review Score:

Approximately 6.8 out of 10

Common Positive Themes:

- The film was frequently described as **"fun,"** **"entertaining,"** and **"quirky,"** with audiences enjoying its humor and nostalgic value.
- Many reviewers found the movie to be a **"blast"** and praised it for its enjoyable and lighthearted approach, highlighting its appeal particularly for younger viewers or longtime fans of the game.

Common Negative Themes:

- Some reviewers described the movie as **"predictable,"** **"boring,"** and **"flawed,"** indicating issues with storytelling and originality.
- Criticism also centered around it being occasionally **"ruined"** by CGI or other production aspects, suggesting technical or visual shortcomings detracted from the overall experience.

Minecraft Movie

Conclusions

- The model is proven very effective at the specific sentiment analysis of movie reviewers, beating out a general purpose LLM at the task

- Analyzing Strong / Weak sentiment is much harder and less accurate than simply an overall positive and negative, but it allows for a more accurate score average and gives more nuance to the summary

- Review length does not seem to be a major factor towards the prediction of sentiment, as the averages for each category were very similar except for the strong negative, which there was not enough data on

- Reviewers who write reviews can have a large bias compared to those who simply leave a number or star rating

Q & A