

Gathering Insights from Movie Reviews with Text Processing and AI

BUAN6390.001 – Analytics Practicum - F25

Group 6

Priyadarshan Parida

Sunayana Jana

Soniya Rajappan

Table of Contents

Introduction.....	3
Methodology.....	3
Data Exploratory Analysis.....	3
Text Preprocessing	3
Traditional Machine Learning Sentiment Analysis.....	4
Generative AI Sentiment Analysis.....	4
Reviewer Segmentation	5
KNIME Data App.....	5
Traditional ML vs Generative AI Models	5
Sentiments and Taste Segments	6
Top and Bottom Movies: Rankings and Insights	6
Key Takeaways.....	7
Limitations	7
Conclusion	7
Appendix	8

Introduction

People rely heavily on movie reviews when choosing what to watch, and streaming platforms use these reviews to improve recommendations and understand audience preferences. Because reviews are written by many different users, they vary widely in length, tone, and detail. This makes it important to analyze them in a structured way.

The goal of this project is to study movie reviews using both traditional machine learning methods and generative AI methods. We compare how each approach understands sentiment, what patterns they find, and how well they identify different types of reviewers. We also use these methods to rank movies based on sentiment and to explore what separates top movies from poorly rated ones.

Methodology

This project was carried out in KNIME using Text Processing nodes, Machine Learning nodes, and OpenAI nodes through the KNIME AI Extension. The process included preparing the reviews, predicting sentiment using two different approaches, grouping reviewers into meaningful segments, and building Data App to display results. The screenshots and visualizations are included in the Appendix.

Data Exploratory Analysis

We loaded the movie review dataset into KNIME and performed basic checks. This included checking for duplicates, missing values, and review lengths. These steps helped us understand data quality before moving forward.

We also identified a data quality issue. The first sentence of every review appeared again at the very end of the same review, resulting in duplicated content. This was consistent across all rows. We corrected this problem using regular expressions (Regex) inside a KNIME String Manipulation node, ensuring each review retained only its true content. This fix prevented misleading patterns in both sentiment scoring and reviewer clustering.

This is Step 1 in the workflow (see Appendix A2).

Text Preprocessing

We used KNIME's Text Processing tools to clean and prepare the reviews for analysis. The main steps included:

- Converting all text to lowercase
- Removing stop words
- Removing punctuation

- Removing numbers
- Applying Snowball Stemming
- Creating Bag-of-Words
- Creating TF-IDF representations

Refer to Appendix A2 for the complete workflow. This segment is Step 2.

Traditional Machine Learning Sentiment Analysis

Using the TF-IDF features, we trained three machine learning models:

- Naïve Bayes
- Support Vector Machine (SVM)
- Logistic Regression

Before training the models, we split the dataset into 70% training data and 30% testing data. The models learned from the training portion and were then evaluated on the test portion. Each model predicted whether a review was positive or negative. We compared their performance using accuracy and confusion matrices. The results are shown in the Appendix A8.

Generative AI Sentiment Analysis

We used the KNIME AI Extension and the LLM Prompter node to generate sentiment using a large language model. This was done in two parts: overall sentiment (Step 4A) and aspect-based sentiment analysis (Step 4B). The detailed workflow is shown in the Appendix A2.

Overall Sentiment Using LLM

In this step, we used the LLM Prompter to classify each review as POS (positive) or NEG (negative). The workflow included:

- OpenAI Authenticator and OpenAI LLM Selector - to select the LLM
- LLM Prompter - where the sentiment prompt was applied
- Rule Engine and GroupBy - to compare predicted labels with true labels
- Expression node - to calculate accuracy

We used a simple instruction prompt that asked the LLM to return only “POS” or “NEG.” This allowed us to directly compare GenAI performance with the machine learning models.

Aspect-Based Sentiment Analysis

For aspect-based sentiment, we analyzed six aspects of each review (acting, story, direction, cinematography, soundtrack, pacing). The workflow used:

- Few-shot examples to show the LLM the expected JSON format
- A combined prompt that inserted each review into the template
- The LLM Prompter, which returned sentiment (positive/neutral/negative) for each aspect
- Validation and JSON extraction using Row Filter, String to JSON, and JSON Path

- A final CSV Writer to store the extracted aspect sentiments

This process gave structured aspect-level insights that traditional ML methods cannot provide.

Reviewer Segmentation

To understand differences in reviewer behavior, we used two methods:

- ML Clustering: We applied k-means to TF-IDF features to form reviewer groups based on word patterns.
- GenAI Personas: We asked the LLM to summarize the main characteristics of each cluster, which produced clear reviewer personas such as “Emotional Reviewers,” “Plot Critics,” and others.

This part of the workflow is identified as Step 5A and 5B in the Appendix A2.

KNIME Data App

We built a KNIME Data App using Component views to present the final insights. The app includes:

- ML and GenAI model comparison dashboard
- Sentiments and Taste Segments dashboard
- Top 10 and Bottom 10 Movies and Genres
- Top 10 and Bottom 10 Movies comparison dashboard

Refer to Step 6 in the workflow in the Appendix A2.

Traditional ML vs Generative AI Models

We compared traditional machine learning models (Logistic Regression, Naïve Bayes, SVM) with the Generative AI model using the same dataset.

The traditional ML models (Naïve Bayes, SVM, Logistic Regression) gave stable and consistent sentiment predictions. Because they rely on TF-IDF features, they responded strongly to clear positive and negative keywords. This made them reliable for straightforward reviews, but they missed subtle tone, context, and mixed feelings.

The generative AI model worked directly with the full text and understood context, tone, sarcasm, and emotional cues. It handled long, narrative-style reviews better than ML and produced more accurate sentiment for reviews with mixed or indirect wording. It also generated explanations and aspect-level sentiment, which ML cannot do.

In terms of accuracy and performance, Generative AI clearly outperformed all traditional ML models. GenAI achieved 90%+ accuracy while the ML models ranged between 48–53% accuracy. The dataset was balanced between positive and negative reviews, so the performance gap was not caused by class imbalance.

From the aspect-based sentiment scores, Acting and Cinematography received the most positive sentiment. Pacing showed strong negative sentiment. Story, Direction, and Soundtrack showed mixed or slightly negative sentiment.

We also built a dashboard in the KNIME Data App to compare all models side by side (Refer to Appendix A8). The dashboard displays accuracy, precision, recall, and F-measure for each model, along with sentiment distribution charts and aspect-level scores. This made it easy to visually confirm the performance gap between ML and Generative AI.

Sentiments and Taste Segments

We used both machine learning and Generative AI to group reviewers based on writing style and sentiment. All segment visuals and comparisons are shown in the Appendix A9.

The ML clusters were uneven in size and harder to interpret because they were based only on TF-IDF word patterns. Sentiment also varied widely across clusters, showing limited consistency.

GenAI produced clear and interpretable personas, such as Entertainment Seekers, Melodrama Buffs, Plot Critics, and Realism Seekers. These groups were more balanced and meaningful because GenAI used full context, tone, and writing style, and not just keywords.

The aspect heatmap in the Data App showed distinct preferences, for example, Melodrama Buffs respond strongly to Acting and Story; Entertainment Seekers prefer positive pacing and overall experience; Plot Critics score lowest across most aspects, especially Pacing.

Overall, ML provided coarse groupings, while GenAI revealed more human-like segments. Sparse data makes the LLM workflow better for extracting meaning than traditional word-frequency methods. Data App dashboards made differences easy to compare through side-by-side charts.

Top and Bottom Movies: Rankings and Insights

We ranked movies using the Wilson Score to account for both sentiment and confidence. This scoring method reduces the impact of small review counts and gives a more reliable ranking. The results are shown in the Top and Bottom Movies Dashboard in the KNIME Data App (Appendix A10).

The top movies had higher Wilson scores (0.57–0.61) and showed broad positive reception across genres. Top movies come from a mix of genres, with several strong Drama titles. The bottom movies had lower scores (~0.51–0.57) and were concentrated in Comedy and Horror.

The top and bottom movies showed clear differences in how people wrote about them. Top movies received longer reviews (~ 1,771 characters on average), while bottom movies had shorter reviews (~ 1,210), often direct. Longer reviews typically contained more emotion, detail, and reasoning.

The Aspect-based scores revealed why some movies performed better. Top movies scored strongly across Acting, Story, and Direction. Bottom movies showed negative sentiment across most aspects, especially Pacing. Every aspect had higher sentiment for the top group. (Appendix A11)

Reviewer composition also differed between the two groups: Top movies attracted a more diverse mix of reviewer personas, while bottom movies were dominated by Plot Critics (86% vs 21%). Top movies also showed far more emotional language (around 42% vs almost 0% for bottom movies).

Key Takeaways

Across all steps of the analysis, several patterns were consistent:

- GenAI outperformed traditional ML, especially for complex, emotional, or sarcastic reviews.
- Aspect-based sentiment helped explain why certain movies were liked or disliked.
- Top movies had richer reviews, more positive aspect sentiment, and more diverse reviewer groups.
- Bottom movies attracted more Plot Critics and showed negative sentiment across nearly every aspect.
- Reviewer personas created by GenAI offered clearer, more human-interpretable segments than ML clusters.

Using multiple methods together, ML for structure and GenAI for depth, gave a broader and more reliable view of sentiment in the dataset.

Limitations

- Genre field was not available in the dataset and was inferred from external lookup.
- Review text varied greatly in length, which can affect both ML and GenAI sentiment.
- GenAI responses can vary slightly based on prompt phrasing.
- Large language models require more compute time and API costs compared to ML.

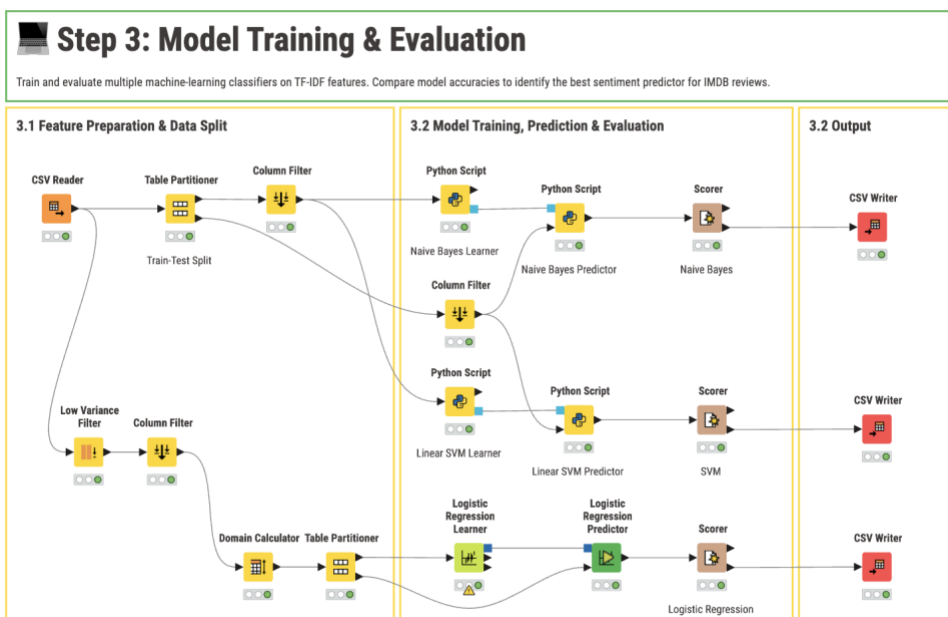
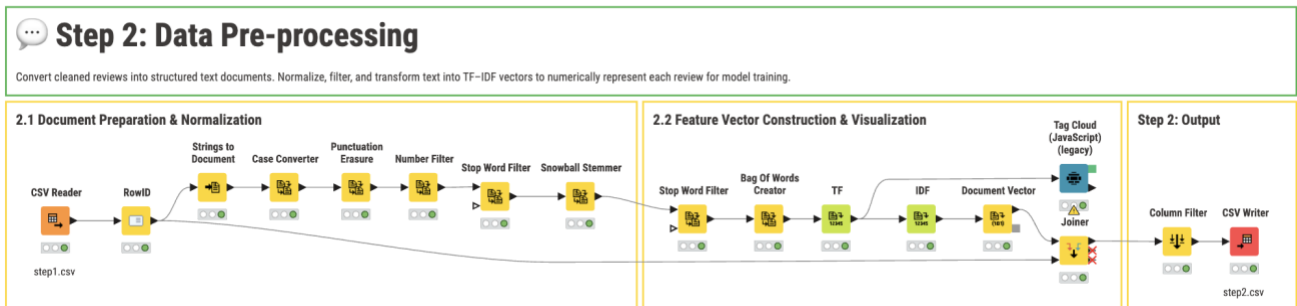
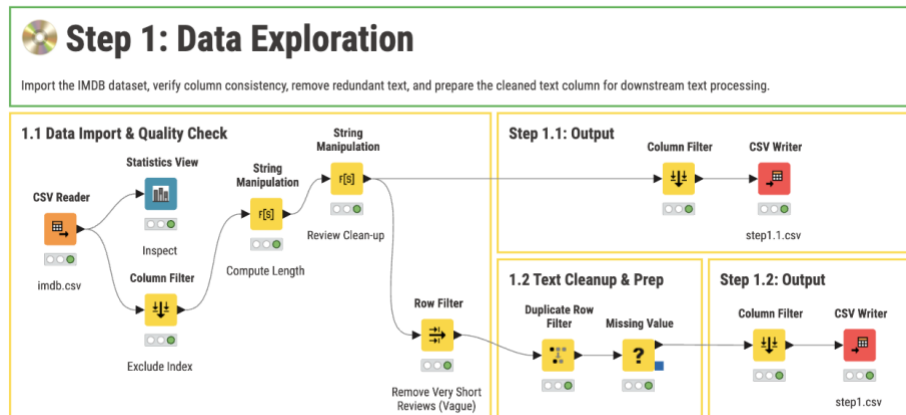
Conclusion

Overall, this project showed that combining text processing, traditional ML, and Generative AI provides strong and complementary insights. GenAI in particular added depth, capturing nuance, explaining sentiment by aspect, and revealing clear reviewer personas. The KNIME workflow and Data App brought all of these insights together in a transparent, reproducible way.

Appendix

A1. Link to the KNIME workflow: <https://github.com/priyadarshanparida/movie-reviews-buan6390>

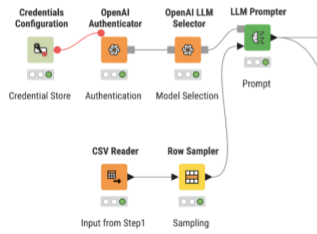
A2. KNIME Workflow screenshots:



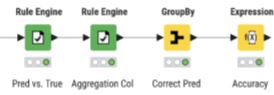
Step 4A: Predict the Overall Sentiment Using Generative AI

Application of generative AI and prompting techniques to predict the overall sentiment of reviews.

4.A.1 LLM Prompter Set-up and Invocation for Sentiment Analysis



4.A.2 Generative AI Model Evaluation



4.A.3 Output



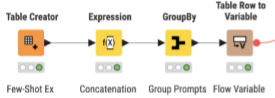
4A Prompt

You are a sentiment analysis expert. For the following movie review text, your job is to classify its sentiment as POS (positive) or NEG (negative). Respond with only one label: POS or NEG.

Step 4B: Generative AI and Prompting Techniques for Aspect Based Sentiment Analysis

Using Generative AI for aspect based sentiment analysis.

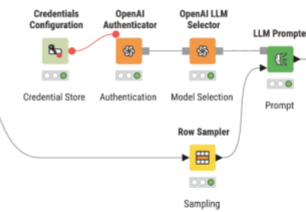
4.B.1 Create Few-Shot Examples Text Block



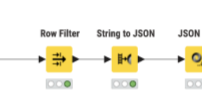
4.B.2 Inject Per Row Prompt



4.B.3 ABSA Using LLM



4.B.4 Validation & Column Extraction



4.B.5 Output



4B Prompt

You are an aspect-based sentiment analysis expert. Analyze the following movie review across six aspects: acting, story, direction, cinematography, soundtrack, and pacing. Use the few-shot examples below for guidance. Few-shot examples: User Review: The reason the DVD releases of this film are inferior and badly handled. Output: ("acting":"neutral","story":"negative","direction":"negative","cinematography":"neutral","soundtrack":"neutral","pacing":"negative") User Review: I'm not a John Cleeze completist although I think he's a genius. The humor and timing were excellent. Output: ("acting":"positive","story":"positive","direction":"positive","cinematography":"neutral","soundtrack":"neutral","pacing":"positive") Now, analyze the user review below and respond ONLY in JSON format as: ("acting":"-","story":"-","direction":"-","cinematography":"-","soundtrack":"-","pacing":"-")
User Review: ...

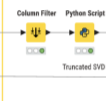
Step 5A: Reviewer Segmentation Using Traditional ML Techniques

Application of traditional ML techniques to segment reviewers based on their taste of movies.

5.A.1 Data Loading



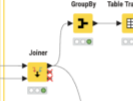
5.A.2 Dimensionality Reduction



5.A.3 K-Means Clustering



5.A.4 Cluster Profiling & Interpretation



5.A.5 Output



5.A.6 Word Cloud



Step 5B: Reviewer Segmentation Using Generative AI Techniques

Application of Generative AI techniques to segment reviewers based on their taste of movies.

5.B.1 Open AI Authentication and LLM Selection



5.B.2 LLM Prompter



5.B.3 Validation & Data Extraction



5.B.4 Raw Clustering Output



5.B.5 Reviewer Taste Segments Summary



5.B.5 Clustering Basis Distribution Summary



5.B.5 Keyword Distribution Summary



5.B.6 Word Cloud



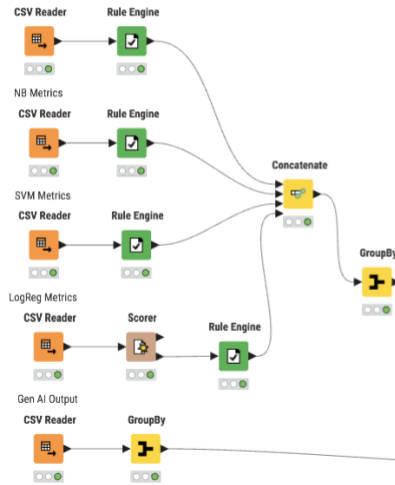
5B LLM Prompt

You are a film-critic analyst.
Read the review and decide what kind of movies the reviewer tends to enjoy.
Classify the review into ONE of these six "Reviewer Taste Segments":
1. "Realism Seekers" - prefer believable stories and authentic emotion.
2. "Plot Critics" - judge films by story logic and writing quality.
3. "Melodrama Buffs" - drawn to emotional or romantic stories.
4. "Entertainment Seekers" - enjoy fun, fast-paced, or action-heavy films.
5. "Genre Loyalists" - consistently favor or dislike specific genres.
6. "Other / Unclassifiable" - no clear or consistent taste signal.
Return JSON in this exact format:
{
 "taste_label": "one of the six segments above",
 "keywords": "[1-3 words that influenced your choice]",
 "rationale": "choose one basis only: tone | emotion | implicit meaning | structure | overall impression | keywords"
}
Guidance:
Choose the option that best reflects what influenced your decision most.
Use 'keywords' only when classification depends mainly on explicit words such as genre names or direct sentiment terms.

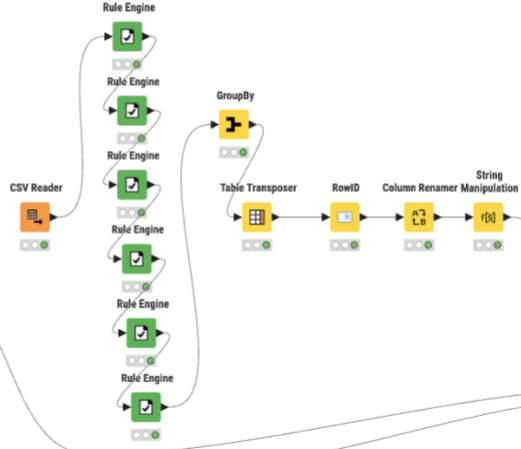
Step 6: Data app for insights

Build dashboards for data visualizations and insights

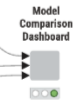
6.A.1 Aggregate Model Comparison



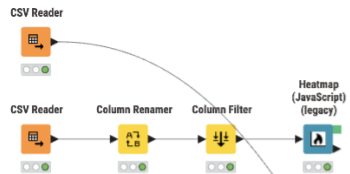
6.A.2 Aspect Based Sentiment



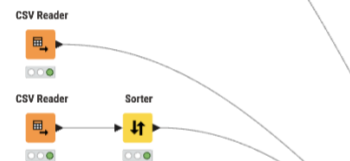
6.A.3 Model Comparison Dashboard in the Data App



6.B.1 Traditional Machine Learning Reviewer Segments



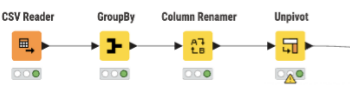
6.B.3 GenAI Reviewer Segments - Cluster Distribution and Cluster Basis



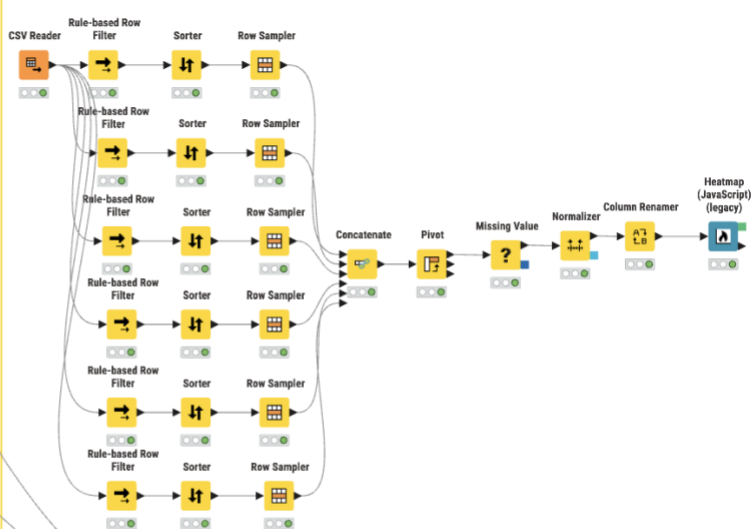
6.B.4 Average Sentiment per ML Cluster



6.B.5 Aspect Sentiment per GenAI Cluster



6.B.2 GenAI Reviewer Segments - Keyword Heatmap



6.B.6 Sentiments and Taste Segments Dashboard in the Data App



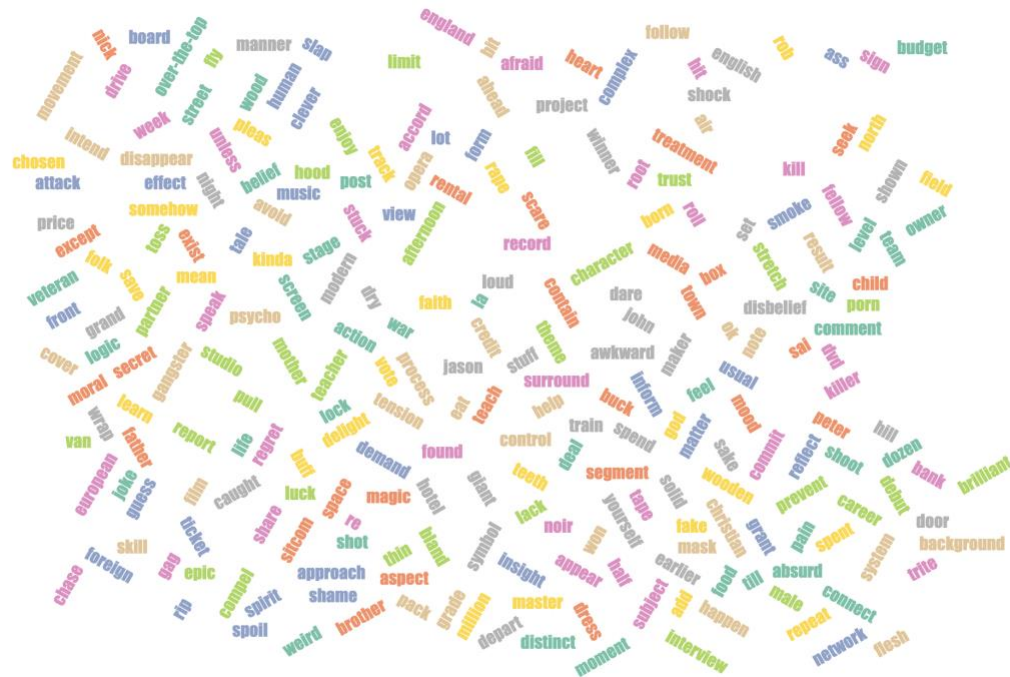
A4. Tag cloud based on TF (Only the first 250 tags are displayed)



A5. Traditional ML clusters

Cluster	Label	High-Weight Words	Interpretation
Cluster 0	Mainstream Enthusiasts	film, watch, time, love, act	Seek balanced enjoyment – acting + story + pace. Represent general audience taste.
Cluster 1	Realism Seekers	real, life, people, character, story	Prefer believable, human stories and emotional authenticity.
Cluster 2	Plot Critics	plot, scene, story, bad, waste	Judge movies on narrative logic & writing quality.
Cluster 3	Melodrama Buffs	recommend, music, perfect, real, laugh	React to emotional and musical resonance; value warmth and connection.
Cluster 4	Entertainment Seekers	fun, laugh, action, recommend, enjoy, watch	Care most about humor, pace, and excitement – the “fun factor.”

A6. Traditional ML clusters – Tag Cloud (Only the first 250 tags are displayed)



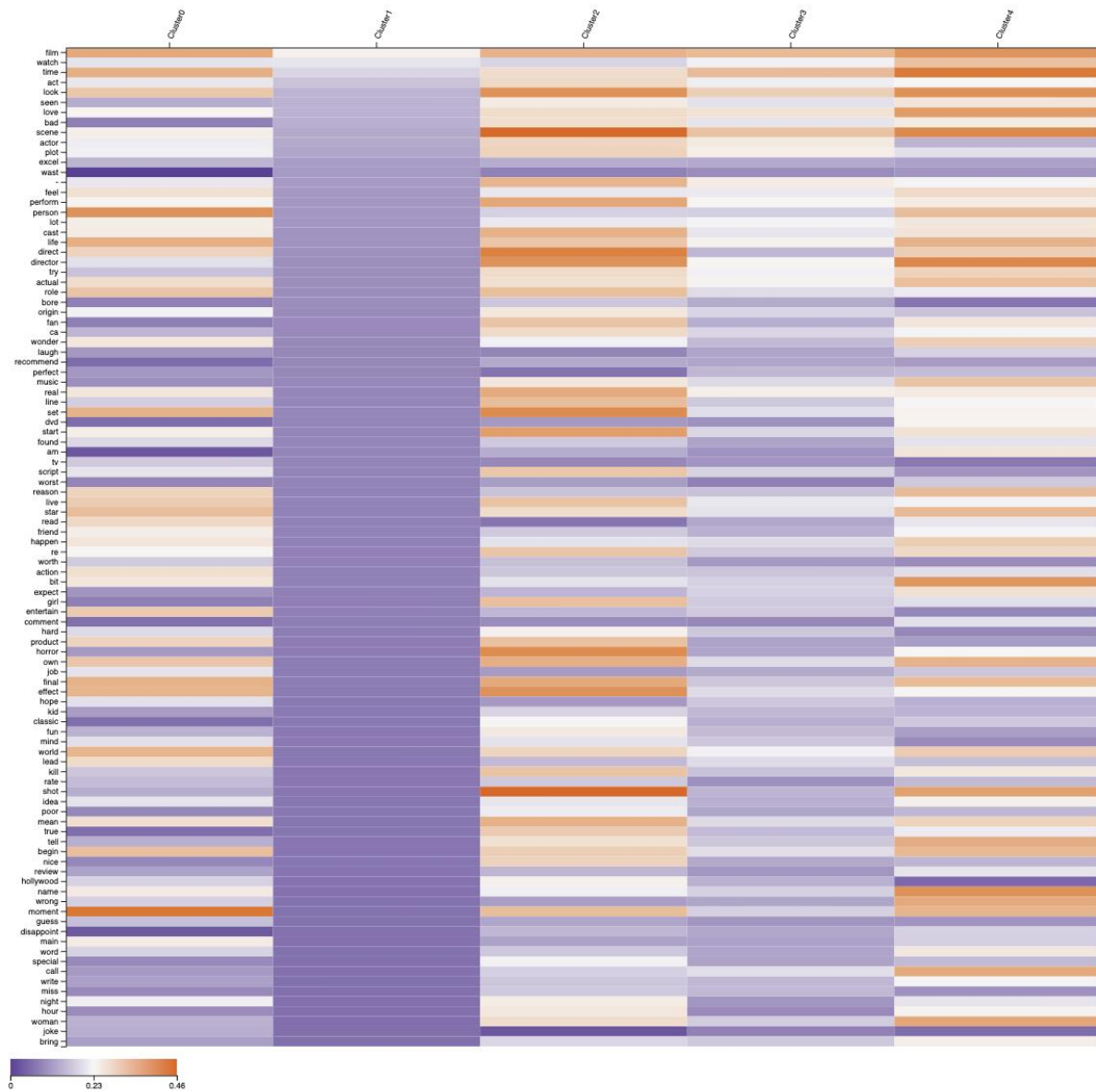
A7. GenAI clusters

Cluster	Label	Top Key Words	Interpretation
1	Entertainment Seekers	fun, entertaining, comedy, hilarious, action, laugh, enjoyable, exciting, humor	Care most about humor, pace, and excitement – the “fun factor.”
2	Genre Loyalists	horror, sci-fi, Star Trek, animation, fantasy, anime, classic, comedy, low budget, martial arts	Consistently praise or criticize specific genres (e.g., horror, comedy, sci-fi, romance).
3	Melodrama Buffs	emotional, romantic, love, emotional stories, romance, love story, touching, beautiful, emotion, heartwarming	Drawn to emotional, romantic, or inspirational films with strong feelings or music.
4	Plot Critics	story logic, writing quality, boring, story, character development, predictable, script, plot, bad acting, characters	Focus on story logic, writing quality, and script coherence.
5	Realism Seekers	authentic emotion, authentic, believable stories, documentary, realistic, believable, realism, true story, emotional, real	Value believable stories, authentic emotion, and human characters.
6	Others	bizarre, disturbing, nostalgia, historical context, interpretation, low budget, psychology, spoof, weird	

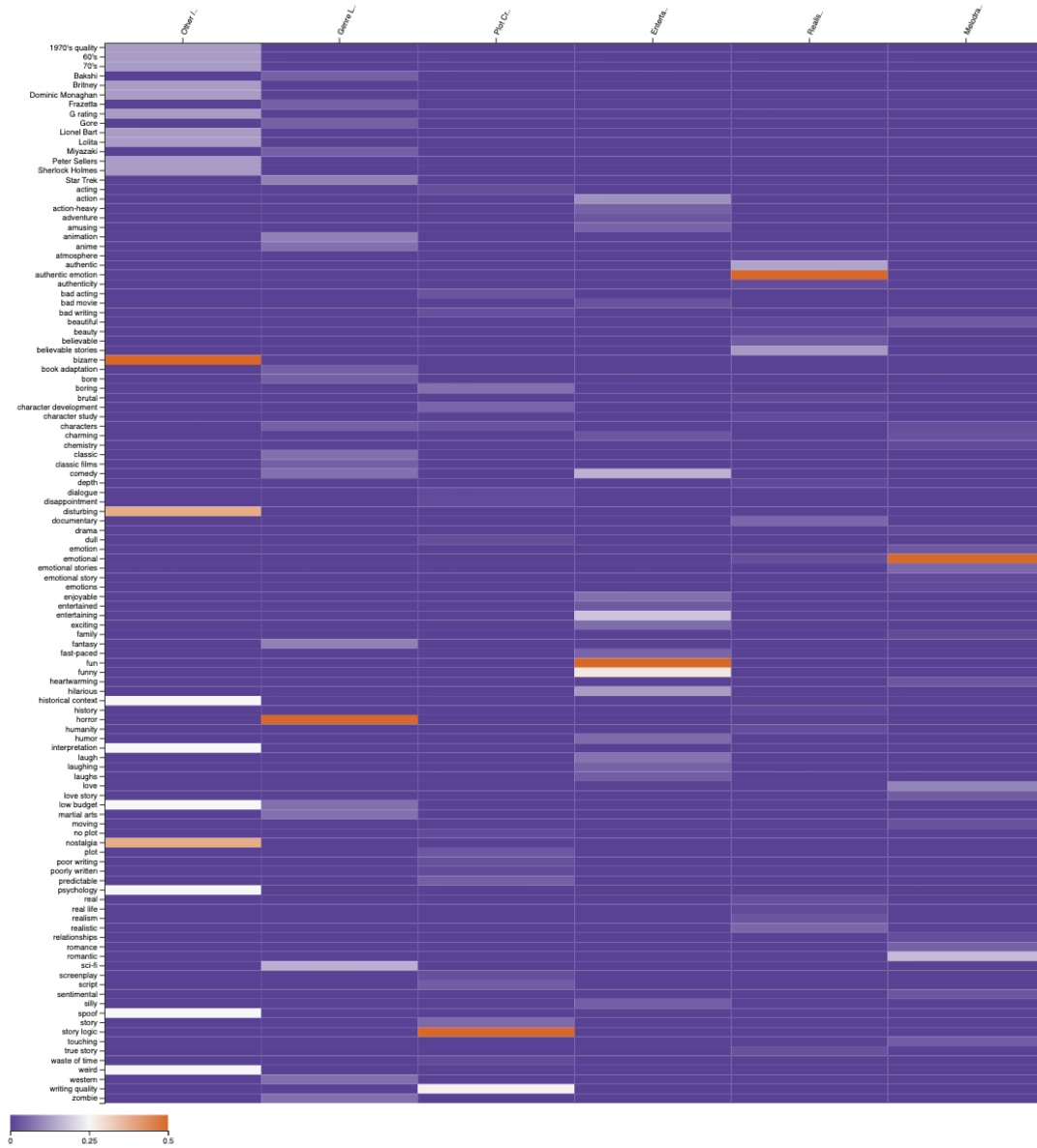
A8. GenAI clusters – Tag Cloud



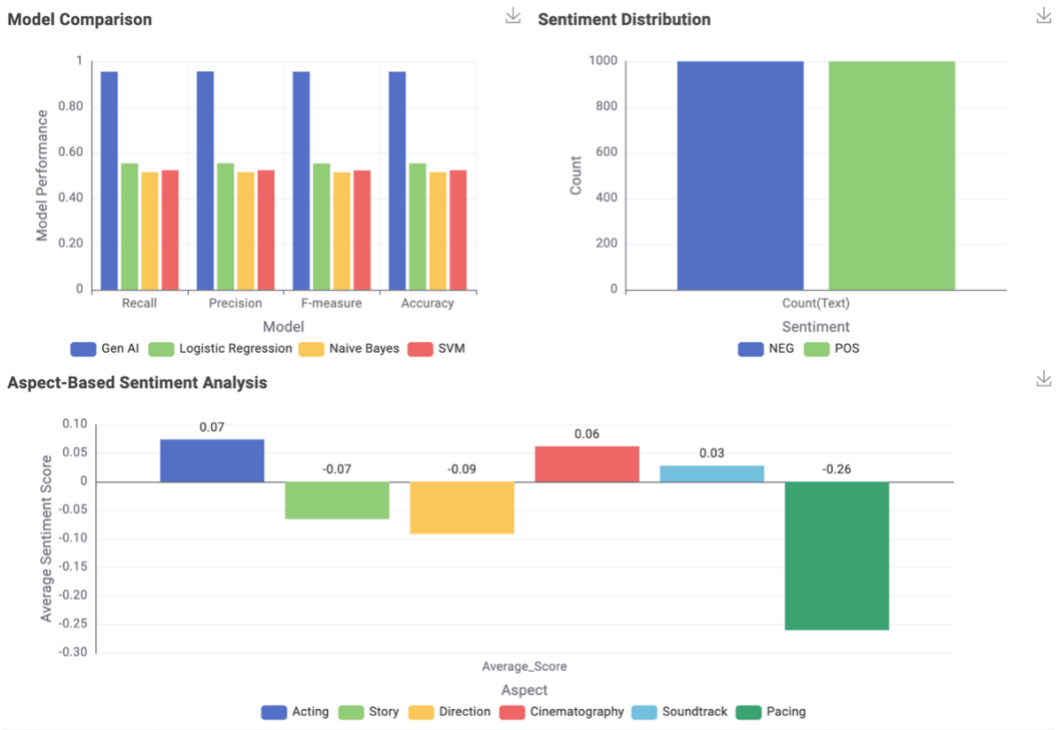
A9. Traditional ML keywords - clusters Heatmap:



A10. GenAI keywords - clusters Heatmap:



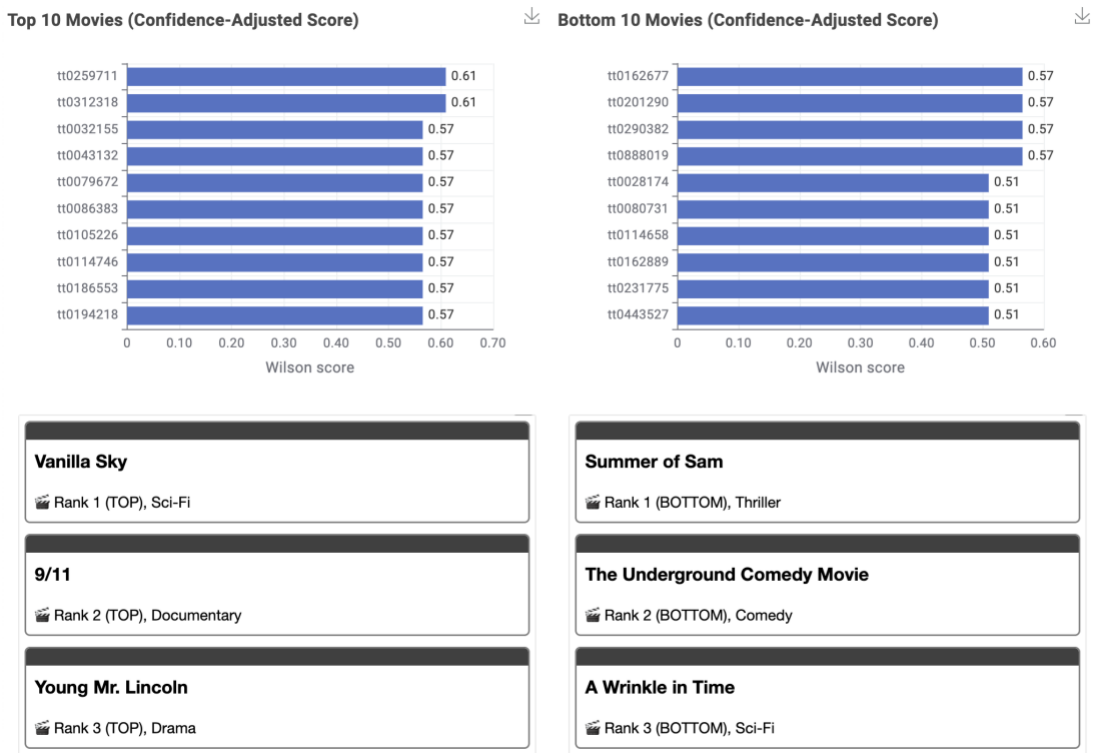
A11. Model Comparison Dashboard:



A12. Sentiments and Taste Segments Dashboard:



A13. Top 10 and Bottom 10 Movies and Genres Dashboard:



A14. Top 10 and Bottom 10 Movies Comparison Dashboard:



Note: the warning signs indicate that there are one or more categories with no values.