

Model Evaluation

A qualitative analysis was performed by comparing outputs from two prompt variations using the meta-llama/Llama-3.1-8B-Instruct model.

Sample Review Analyzed

'brings,back,,original,horror,,73'

Qualitative Analysis of Prompt Outputs

Result from Prompt 1 (Direct Template)

Output: "The sentiment of this review is likely: **Positive**... The reviewer mentions 'original horror'... The use of a numerical rating of 73 also suggests a high rating..."

- **Analysis:** The model correctly identified the sentiment as **Positive**. It understood the key phrase "brings back original horror" but made a minor error, misinterpreting the year '73 as a rating.

Result from Prompt 2 (Role-Playing Template)

Output: "I'd say it's a **bad review**... The use of multiple commas...suggests...uncertainty or frustration."

- **Analysis:** The model **failed**, classifying the review as **Negative**. The persona prompt caused the model to over-analyze and "hallucinate" incorrect logic based on punctuation, leading to a wrong conclusion.

Result from Prompt 3 (Few-Shot):

Output: It seems like the review is referring to a movie, likely "Brings Back Original Horror" (assuming "73" is a rating out of 100). However, the text is incomplete and appears to be truncated.

- **Analysis:** That being said, I'll attempt to classify the sentiment based on the available information. Since the review seems to be making a positive statement about the movie's return to original horror, I would classify the sentiment as: **POSITIVE**

Conclusion

For this task, the simpler, direct prompt was more effective. The complexity of the role-playing prompt confused the model, showing that elaborate instructions can introduce error. However, the 3rd prompt analyzed the semantic meaning of the review but was incorrect with the movie name and interpreted 73 as the rating instead of the year.

Troubleshooting

Identified Issue: Low Available Data

The small dataset size (220 reviews) presented a challenge, which can lead to underfitting, bias, and poor generalization.

Proposed Solution: Generating Synthetic Text with Faker

The **Faker library** can be used to augment the dataset. It generates new, plausible-looking text data from templates (e.g., "{adjective} movie with {adjective} acting."). This increases the volume and diversity of the training data, helping to mitigate the risks associated with a small dataset.