

MovieMoods: DSA 495 Final Project Report

Name: Nitesh Kanamarlapudi

Class: DSA 495 - Fundamentals of Large Language Models

Project Repository: <https://github.com/noteesh/MovieMoods>

Date: December 01, 2025

PROJECT OVERVIEW

MovieMoods is a project that combines Fine-tuning, BERT Classification, and RAG to create a movie review retrieval system that is sentiment-aware. MovieMoods classifies user queries by sentiment [positive / negative], routes them to sentiment-matched documents, and retrieves similar movie reviews using embeddings. This project combines 2 core LLM techniques that we learned in DSA 495: Classification/Routing and Retrieval-Augmented Generation (RAG).

Architecture:

1. User Query
2. Sentiment Classifier (DistilBERT)
3. Routing
4. RAG Retrieval (sentence transformers)
5. Results

Key Results & Resources:

- **Classification accuracy:** 95% on 100-sample test set
- **Retrieval Recall@3:** 100% (all test queries found relevant documents)
- **Routing improvement:** +20% average similarity score vs. no routing
- **Training data:** 500 IMDB reviews
- **RAG corpus:** 20 curated movie summaries (10 positive, 10 negative)

IMPLEMENTATION

For MovieMoods's sentiment classification, I used DistilBERT fine-tuned on 500 IMDB movie reviews for three full epochs. I chose DistilBERT because it was more efficient (40% smaller & 60% faster than BERT) while maintaining 97% of BERT's performance. The RAG part of this project uses sentence transformers (all-MiniLM-L6-v2) for dense embeddings and cosine similarity for the retrieval.

I conducted two ablation studies to validate these design choices:

1. Varying top_k (1, 3, 5) showed k=3 balances relevance and diversity (cell 8 in project ipynb)
2. Comparing routing vs. no routing demonstrated 20% improvement in similarity scores when filtering by sentiment before retrieval (cell 9 in project ipynb).

INTENDED USE

MovieMoods is a tool that is designed for movie production teams to decide whether the reviews that they receive on movies are intended to serve as positive or negative. It can also serve to help students and educators as an educational demonstration of how to combine various LLM techniques. It demonstrates how classification can improve retrieval systems through routing. This makes MovieMoods helpful for learning environments and movie production teams that want to classify their reviews.

What problems does MovieMoods solve?

1. Allows movie production teams to evaluate the sentiments of their reviews
2. Demonstrates practical implementation of Router + RAG architecture
3. Shows how sentiment classification improves retrieval relevance
4. Provides a reproducible example on free hardware (Google Colab T4)
5. Offers a template for building sentiment-aware search systems

Appropriate uses:

- Understanding review sentiments on a movie production team.
- Educational demonstrations for learning NLP concepts.

Inappropriate uses:

- High-stakes decisions.
- Non-movie domains without retraining.
- Systems that require a nuanced sentiment (neutral or based on other parameters).

LIMITATIONS

Biases:

- The IMDB dataset contains demographic biases (skews male, Western, younger reviewers)
- Extreme Ratings (mostly very positive or very negative)
- Genre biases (popular genres are over-represented)
- The 20-document corpus is too small for production use and contains only well-known films, limiting diversity.

Model Limitations:

- The classifier fails on sarcasm ("so bad it's good"), negation ("not the worst"), and mixed sentiment ("great acting but terrible plot").
- The system only handles binary sentiment and can't process neutral sentiments or sentiments based on other parameters.
- The small corpus size limits the retrieval.

Edge Cases:

- Double negatives in the reviews (positive sentiment) confuse the classifier in some cases.

- Queries that are out of the domain (example: “How’s the weather today?”) that aren’t related to the context still get processed, which sometimes produces nonsense results.

SAFETY CONSIDERATIONS

Risk Level:

Low. MovieMoods is a low-risk entertainment tool with no financial, medical, legal, or safety implications. Incorrect recommendations may waste a user’s time but cause no serious harm.

Disclaimers:

1. NOT FOR PRODUCTION: This is an educational demonstration. Do not deploy without expanding the corpus (1000+ documents), testing for bias, implementing error handling, and conducting user safety testing.
2. NOT PROFESSIONAL ADVICE: The system provides movie suggestions only and does not replace professional critics, personalized recommendation systems, or human judgment about content appropriateness.
3. LIMITED SCOPE: Works only for movie reviews in English with binary sentiment [positive / negative]. Not suitable for other domains.
4. POTENTIAL BIASES: May reflect IMDB demographic biases and has limitations associated with a small training dataset.

Content Safety: All 20 corpus documents were manually reviewed to exclude offensive material. All movies rated PG-13 or below.

DATA RIGHTS

IMDB Movie Review Dataset: Source: Stanford University via Hugging Face datasets library
 Citation: Maas, A. L., et al. (2011). Learning word vectors for sentiment analysis. ACL-HLT 2011.
 License: Public domain for academic use
 Privacy: No PII/PHI, only movie review text
 Access: `datasets.load_dataset("imdb")`

Models: DistilBERT: Sanh, V., et al. (2019). DistilBERT, a distilled version of BERT. arXiv:1910.01108. License: Apache 2.0 Sentence-Transformers: Reimers, N., & Gurevych, I. (2019). Sentence-BERT. EMNLP-IJCNLP 2019. License: Apache 2.0

Software Libraries: All use permissive open-source licenses (PyTorch: BSD, Hugging Face: Apache 2.0, scikit-learn: BSD 3-Clause)

All components comply with licenses and are free for educational use.

TRANSPARENCY

LLM Assistance Used:

1. Corpus Generation: The 20 movie summaries were initially generated using Claude (Anthropic). Each summary was then manually validated by cross-referencing with IMDB and Rotten Tomatoes for factual accuracy, verifying sentiment matches critical consensus, and

checking for appropriate length and clarity. 100% of summaries were human-validated before inclusion.

2. Documentation: LLM helped with README formatting and report structure.

Validation Methods:

1. Corpus: I cross-referenced the generated corpus with IMDB/Rotten Tomatoes for accuracy on the sentiment.
2. Documentation: I used the formatting tools, and manually wrote & verified that all the information suggested to include in the formatting was correct and followed the project guidelines.

CONCLUSION

MovieMoods was able to successfully use Routing & Retrieval-Augmented Generation by combining sentiment classifications with RAG. The project was able to achieve a 95% classification accuracy and showed that sentiment-based routing improved retrieval relevance by 20%.

Some of the key findings were that simple routing mechanisms significantly improved retrieval when the sentiment mattered, and DistilBERT was an amazing balance between performance and efficiency for an education application/project like this one.

Lastly, while this system had a few limitations preventing a large-scale production, it was able to successfully show how classification and RAG techniques combine in practice. This project/results are fully reproducible, well-documented, and gave me a foundation for building more sentiment-aware applications in the future.

REFERENCES

1. Devlin, J., et al. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805.
2. Sanh, V., et al. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv:1910.01108.
3. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. EMNLP-IJCNLP 2019.
4. Maas, A. L., et al. (2011). Learning word vectors for sentiment analysis. ACL-HLT 2011.
5. Lewis, P., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. NeurIPS 2020.