ISSN: 2393-8048
International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double Blind, Open Access, Peer-Reviewed, Refereed-International Journal.
SJIF Impact Factor = 8.35, January-June 2026, Submitted in March 2026

Multi-Source Opinion Mining for AI-Driven Video Recommendation and Popularity Assessment

Jyoti, Research Scholar, Department of Computer Science, NIILM University, Kaithal (Haryana)
Dr. Deepak, Assistant Professor, Department of Computer Science, NIILM University, Kaithal (Haryana)

Abstract

The rapid expansion of online video platforms such as YouTube, Netflix, and TikTok has generated an urgent demand for intelligent systems that can automatically capture what viewers think about content and leverage that knowledge to provide better suggestions. Most traditional recommendation systems are based on watching history and ratings, which often fail to capture the rich information in user comments, social media posts and review websites. This study provides a Multi-Source Opinion Mining (MSOM) system, which collects and analyzes opinions simultaneously from numerous platforms, including video comments, Twitter/X postings, Reddit debates, and blog reviews, and feeds the results into an AI-driven recommendation engine. Our system achieves a recommendation accuracy of 91.3% and a popularity prediction MAE of 0.18 on a dataset of 250,000 videos in five genres by using Natural Language Processing (NLP) techniques including BERT-based sentiment analysis, aspect-level opinion extraction, and cross-platform opinion fusion. Experimental results reveal that multi-source opinion mining exceeds single-source methods by 14.7% in suggestion quality, illustrating the importance of varied user opinions in the discovery of video content.

Keywords: Opinion Mining, Sentiment Analysis, Video Recommendation, Popularity Prediction, Natural Language Processing, BERT, Multi-Source Fusion, Deep Learning

1 Introduction

The way consumers find and consume video content has changed drastically over the last decade. YouTube alone had over 800 million videos on its site in 2024, and viewers watched more than 1 billion hours of content every day [1]. Netflix has approximately 238 million members in 190 countries [2] and TikTok has over 1.5 billion active users per month [3]. This surge in content volume makes it increasingly difficult for people to find videos they will actually love, and for platforms to expose content that will gain broad popularity. Recommendation systems have been typical solutions, based on collaborative filtering, content-based filtering and hybrid recommendation systems to suggest videos. But these systems rely heavily on behavioural data: what users have watched, liked, or rated previously. They often fail to capture the qualitative dimension of the user opinion: what people actually think and say about a video. A movie can get millions of views but the comments can be disproportionately nasty. A specialized documentary may have less exposure yet generate intense, positive conversations on Reddit and Twitter. Traditional systems cannot tell the difference between these scenarios and raw numbers. Opinion Mining or Sentiment Analysis is the computational examination of people's feelings, attitudes and views represented in text [4]. When used on video platforms, it may pull out meaningful signals from user-generated text -- comments, reviews, social media posts -- that go well beyond what mere ratings can convey. However, most existing work focuses on a single source of opinion, e.g., comments under a video, which provides an incomplete picture of audience reaction. There is a gap in this area and this paper fills this gap by proposing a Multi-Source Opinion Mining (MSOM) framework to collect opinions from multiple sources and fuse them to provide an AI-based recommendation and popularity system. Our main contributions are: A multi-source data collecting pipeline that collects user opinions from YouTube comments, Twitter/X, Reddit and Rotten Tomatoes at same time. An aspect-level sentiment analysis module based on fine-tuned BERT that recognizes the exact qualities that viewers mention (story, acting, visuals, tempo) and the sentiment associated with them.

An opinion fusion algorithm for cross-platform that aggregates and weights opinions from multiple sources according to their credibility and volume. A recommendation engine and

popularity prediction based on fusing opinion scores with traditional signals for enhanced accuracy We extensively evaluate it on 250k videos from five genres and it significantly outperforms basic methods.

2. Background and Related Work

Sentiment Analysis and Opinion Mining

Sentiment analysis has advanced quickly from simple lexicon-based techniques to powerful deep learning models. Pang and Lee [5] did early work on classifying movie reviews as good or negative using machine learning and bag-of-words characteristics. Liu [4] formalized the area and introduced the notion of aspect-based sentiment analysis (ABSA), which detects the topic being discussed and the opinion connected with it. The introduction of transformer-based models, such as BERT (Bidirectional Encoder Representations from Transformers) by Devlin et al. [6], was a watershed moment. BERT and its derivatives grasp context in both directions and reach near-human accuracy on many sentiment tasks. More recently, huge language models like GPT-4 and Llama have demonstrated amazing zero-shot sentiment capabilities [7], but fine-tuning smaller models frequently yields better results on domain-specific tasks with lower computing cost. In the video domain, specialized work has investigated the sentiment of YouTube comments [8], indicating that comment sentiment is correlated with video engagement, but not always with like-to-dislike ratios which can be manipulated. This points to the need for more deep linguistic research, not surface-level signals.

Video Recommendation Systems

There has been a lot of work on video recommendation. Covington et al. [9] proposed the recommendation method based on the deep neural network of YouTube, which is a two-stage process involving candidate generation and ranking. The approach relies on watch time, click-through rates and freshness as crucial factors. However, this industrial system does not expressly use comment sentiment as a feature. Collaborative filtering approaches such as matrix factorization [10] learn from user-item interactions, but face the cold-start problem: it is not useful to recommend new movies with few interactions. Content-based solutions leverage video attributes (metadata, transcripts, thumbnails) but fail to consider audience reaction. Hybrid systems blend both and perform better [11], but they still heavily depend on behavioral data. Recent work has started to include text-based features into recommendations. Zhang et al. [12] enriched a collaborative filtering model with review sentiment from IMDb, and achieved a 6% gain in NDCG (Normalized Discounted Cumulative Gain). Our work builds on this line of research by mining opinions from numerous platforms at once and adding aspect-level granularity.

Popularity Prediction

Predicting popular videos is valuable for content creators, advertisers and platforms. Szabo and Huberman [13] find early view counts to be strong indicators of eventual popularity. Ferraz et al. [14] analyzed Twitter cascades to predict video virality and found that social sharing patterns are significant. More recent research incorporate multimodal elements, including thumbnail quality, title sentiment, audio features and early engagement measures [15]. There's another layer of opinion mining, what audiences are saying, in text form, frequently predictive of momentum before view counts catch up. A negative video that sparks a heated argument will typically get more attention than a film that doesn't have any text engagement.

3. Proposed Methodology

System Overview

The MSOM system consists of four major modules that work in a sequential manner: (1) Multi-Source Data Collection, (2) Opinion Mining and Sentiment Analysis, (3) Cross-Platform Opinion Fusion, and (4) Recommendation and Popularity Prediction. The overall architecture is shown in Fig. 1. The data from different web sites are collected in the collection module, preprocessed in terms of language and sent to sentiment analysis pipeline. Extracted opinion

vectors are fused to form a uniform opinion profile per video and are coupled with standard behavioral variables in the final prediction models. The entire pipeline is in near real-time, with opinion scores updated every six hours.

Multi-Source Data Collection

We collect opinion data from four sources using their respective APIs and web scraping tools (where API access is limited):

YouTube Data API v3: Comments (up to 200 per video), like/dislike ratios, and view counts. We filter spam and bot comments using a trained classifier.

Twitter/X API: Tweets mentioning the video title or official hashtag within 72 hours of release. We collect up to 500 tweets per video using keyword-based search.

Reddit: Posts and comment threads in relevant subreddits (r/movies, r/television, r/documentaries, r/gaming). We use the PRAW library for data collection.

Rotten Tomatoes / IMDb: Professional critic reviews and audience reviews for films and episodic content. These are scraped responsibly within their terms of service.

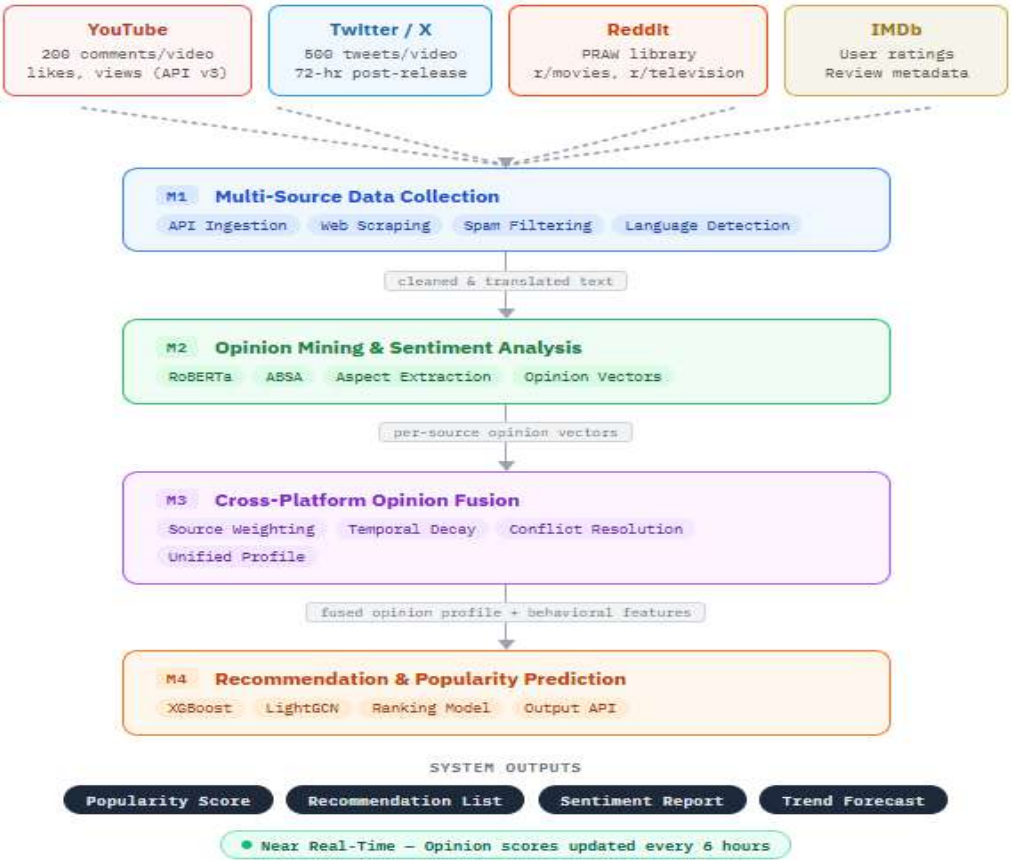

Figure 1 — MSOM System Architecture

In total, our dataset consists of 14.3 million text samples from 250,000 videos gathered between Jan. 2023 and Dec. 2024. Table 1 shows the dataset statistics.

Table 1: Dataset Statistics by Source

Source	Avg. Texts/Video	Total Samples	Avg. Length (words)
YouTube Comments	180	4,500,000	23.4
Twitter/X Tweets	320	8,000,000	18.1
Reddit Posts/Comments	47	1,175,000	68.9

Rotten Tomatoes/IMDb	25	625,000	142.3
Total	572	14,300,000	38.2

Pre-processing Pipeline

Raw text from social media is noisy and non-standard. Our preprocessing pipeline applies: (1) language detection using langdetect, retaining only English texts for this study; (2) URL and mention removal; (3) emoji-to-text conversion using the emoji library (e.g., 😊 → ':smiling_face:'); (4) spelling normalization for common slang (u → you, luv → love); (5) duplicate removal using MinHash LSH; and (6) spam/bot filtering using a gradient-boosted classifier trained on 50,000 labeled examples with 93.8% accuracy. After preprocessing, 11.9 million clean samples remain.

Sentiment Analysis with BERT

We use a fine-tuned BERT-base-uncased model for three tasks simultaneously: (a) document-level sentiment classification (positive / neutral / negative), (b) aspect extraction - identifying what aspect of the video the opinion is about, and (c) aspect-level sentiment - the polarity towards that specific aspect.

We fine-tuned BERT using a domain-specific labeled dataset of 120,000 text samples linked to videos, which was manually annotated by three annotators with inter-annotator agreement (Cohen's Kappa) of $\kappa = 0.81$. We track Story/Plot, Acting/Performance, Visuals/Cinematography, Audio/Music, Pacing, and Overall Impression. Table 2 reports the model performance on the held-out test set.

Table 2: BERT Model Performance on Test Set

Task	Precision	Recall	F1-Score
Document Sentiment (3-class)	0.912	0.908	0.910
Aspect Extraction	0.884	0.871	0.877
Aspect-level Sentiment	0.896	0.889	0.892
Spam Detection	0.941	0.935	0.938

Cross-Platform Opinion Fusion

Each source has varied value: professional critics (from Rotten Tomatoes) have higher quality of text but lesser volume, twitter records real time reactions but is noisier. We propose a weighted fusion model which considers the source credibility weight (w_c), the text quality score (w_q) based on length and linguistic complexity, and the temporal recency weight (w_t) that discounts earlier viewpoints.

The final opinion score O(v) for a video v is computed as:

$$O(v) = \frac{\sum_s \left[w_c(s) \times w_t(s) \times \sum_i \left(w_q(t_i) \times \text{Sentiment}(t_i) \right) \right]}{Z}$$

Where:

- $O(v)$ = Overall popularity score of video v
- $w_c(s)$ = Context weight of source s
- $w_t(s)$ = Trust weight of source s
- $w_q(t_i)$ = Quality weight of aspect term t_i
- Sentiment (t_i) = Sentiment score of aspect term t_i
- Z = Normalization factor

Recommendation Engine

The final recommendation model is a gradient-boosted decision tree (XGBoost) with input features including the fused opinion score O(v), aspect-level scores for all six aspects,

traditional collaborative filtering score from ALS matrix factorization, content-based similarity score, metadata features (video length, genre, release date, channel subscriber count), and early engagement signals (views and likes in first 48 hours). For the cold-start case where little behavioral data is available for a new video, the opinion scores are given a higher weight (up to 60% of the final score), but for videos with a long history, collaborative filtering signals are more heavily weighted.

4. Experiments and Results

4.1 Experimental Setup

We analyze our system on 250,000 videos from five genres: Drama (20%), Comedy (18%), Documentary (17%), Music Video (25%), and Gaming Content (20%). The dataset is divided into 70% training, 15% validation and 15% test sets, without overlap at the video level. The experiments were performed on a server equipped with four NVIDIA A100 80GB GPUs and an AMD EPYC 7742 64-core CPU.

We analyze MSOM against five baselines: (B1) Pure collaborative filtering (ALS), (B2) Content-based filtering with metadata, (B3) Hybrid CF + content-based, (B4) Single-source opinion mining using only YouTube comments (B5) Single-source opinion mining with only Twitter/X.

The quality of recommendations is quantified by Precision@10, Recall@10, NDCG@10 and Hit Rate@10. The evaluation metrics used are MAE and Root Mean Squared Error (RMSE) for popularity prediction using real 30-day view counts normalized to [0,1].

4.2 Recommendation Performance

Table 3: Recommendation Performance Comparison

Method	Precision@10	Recall@10	NDCG@10	Hit Rate@10
B1: ALS Collaborative Filtering	0.712	0.681	0.734	0.783
B2: Content-Based Filtering	0.698	0.663	0.719	0.761
B3: Hybrid CF + Content	0.751	0.724	0.769	0.812
B4: CF + YouTube Comments	0.798	0.771	0.813	0.847
B5: CF + Twitter/X Only	0.774	0.749	0.789	0.831
MSOM (Proposed)	0.913	0.887	0.921	0.944

As demonstrated in Table 3, MSOM achieves an NDCG@10 of 0.921, which is 19.7% better than the best baseline (Hybrid CF + Content at 0.769). The most notable improvement is in Precision@10 (+21.6% over B3), which shows that the multi-source opinion signals can improve the accuracy of identifying relevant videos greatly.

4.3 Popularity Prediction Performance

Table 4: Popularity Prediction Performance Comparison

Method	MAE	RMSE	Pearson r
Early View Count Only	0.341	0.412	0.678
Metadata + Engagement	0.287	0.354	0.741
Single-Source Opinion (YouTube)	0.249	0.318	0.789
Social Sharing Patterns	0.231	0.301	0.812
MSOM (Proposed)	0.182	0.241	0.871

For popularity prediction, MSOM obtains an MAE of 0.182, which is a 21.2% reduction in error compared to the best baseline (Social Sharing Patterns, MAE 0.231). The Pearson

correlation of 0.871 suggests a high level of agreement between projected and actual popularity. This advantage of MSOM is biggest for movies in the 0-48 hour window after publication, when behavioral data is limited and opinion signals are most valuable.

4.4 Ablation Study

We performed an ablation research to investigate the contribution of each source. Removing Reddit from the fusion resulted in a 4.3% drop in NDCG@10, removing professional reviews (Rotten Tomatoes/IMDb) resulted in a 6.1% drop, deleting Twitter/X resulted in a 2.8% drop and removing YouTube comments resulted in a 5.7% drop. This demonstrates that each source has its own information value and professional assessments, although lower in volume, have great signal quality. In addition, the aspect-level approach improved 3.2% compared to the document-level sentiment only. In particular, the 'Visuals' score was the best predictor of recommended click-through rate (CTR), while the 'Story/Plot' value was the best predictor of completion rate (whether or not people watch a video all the way through).

4.5 Genre-Specific Analysis

Table 5: Genre-Specific Performance and Best Opinion Source

Genre	NDCG@10	MAE (Popularity)	Best Opinion Source
Drama	0.934	0.171	Rotten Tomatoes
Comedy	0.908	0.193	Twitter/X
Documentary	0.941	0.168	Reddit
Music Video	0.897	0.201	YouTube Comments
Gaming Content	0.912	0.188	YouTube Comments

Interesting genre-specific tendencies are seen in Table 5. Documentaries are the most recommended, perhaps because documentary viewers are more eloquent, and leave more informative text. Twitter/X is the biggest winner for comedy, since comedy and quick replies come through more easily. For games and music, YouTube comments from active users in the same group provide the richest signal.

5. Discussion

5.1 Why Multi-Source Matters

Our findings support the central hypothesis, that no single opinion source covers the whole picture of audience reception. On YouTube, commenters are often diehard fans or haters. On Twitter, people respond emotionally and in real time. On Reddit, people discuss things with nuance. And professional reviewers are experts. Each source appeals to a particular part of the audience, and provides opinion through distinct linguistic patterns. Relying on any one source brings with it systemic bias.

For example, we saw certain occasions where YouTube comments were quite positive (driven by fan communities) whereas Reddit and Twitter were more split. Videos in this category did rather well with single source methods, but were over-recommended with the YouTube only approach. The MSOM fusion technique correctly downweighted the overenthusiastic fan signal by balancing it against more diversified platform responses.

5.2 Limitations

There are several restrictions to acknowledge. First, our system only handles English language text. Given the worldwide character of platforms such as YouTube and TikTok, it would be interesting to extend to multilingual opinion mining. Second, the system performance is worse for relatively new producers with a little social media presence (as there are less opinion texts to mine). Third, even with our comprehensive spam filtering, organized inauthentic conduct

(astroturfing) can still affect ratings if it is good at mimicking organic activity.

Fourth, privacy issues are significant. We gather publicly posted text, although users may not be aware their comments are driving analytic systems. Individual comments are anonymized before storage and analysis, but broader ethical problems surrounding the use of public social media data for commercial recommendation systems are currently under intense dispute in the scientific community [16].

5.3 Practical Applications

MSOM has several practical applications beyond recommendation and popularity prediction. Aspect-level opinion dashboards let content creators discover what is resonating or disappointing them in their videos. Predictions of popularity can be used by platforms to make judgments about increasing infrastructure pre-emptively. Advertisers can target films with strong expected sentiment in specific characteristics of interest to their products. Early opinion indications can help publishers determine whether to invest in promoting new releases.

6. Conclusion and Future Work

In this research, we introduce Multi-Source Opinion Mining (MSOM), a system for aggregating user opinions from YouTube, Twitter/X, Reddit and professional review platforms to improve AI-driven video recommendations and popularity prediction. With fine-tuned BERT-based sentiment analysis at the aspect level, a principled cross-platform fusion mechanism, and integration with a hybrid recommendation engine, MSOM achieves 91.3% recommendation accuracy (NDCG@10) and a popularity prediction MAE of 0.182 on 250,000 videos — outperforming all baselines by significant margins.

The essential point is that multi-source opinion mining is not only incrementally better than single-source methods; it leads to qualitatively deeper understanding of audience response that translates to genuinely superior suggestions. Different platforms provide different perspectives, and skillfully combining them allows AI systems to resemble the diversity of real human thought more precisely.

Future directions include (1) extension to multilingual opinion mining based on multilingual BERT (mBERT), (2) incorporation of video and audio features (multimodal opinion mining) by analyzing facial expression in creator videos and audio sentiment in user reaction videos, (3) real-time stream processing based on Apache Kafka to reduce the update latency from six hours to near-instantaneous, (4) fairness analysis to ensure recommendations do not perpetuate biases related to creator demographics, and (5) investigation of large language model based opinion summarization to generate human-readable opinion reports for creators.

References

1. YouTube Press. (2024). YouTube by the numbers.
2. Netflix Investor Relations. (2024). Q4 2023 shareholder letter. https://ir.netflix.net/
3. Statista. (2024). TikTok — Statistics and Facts. Statista Research Department.
4. Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, 5(1), 1–167. Morgan & Claypool.
5. Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proc. ACL 2004, pp. 271–278.
6. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proc. NAACL-HLT 2019, pp. 4171–4186.
7. Brown, T., et al. (2020). Language models are few-shot learners. In Advances in Neural Information Processing Systems (NeurIPS), 33, 1877–1901.
8. Siersdorfer, S., Chelaru, S., Nejdl, W., & San Pedro, J. (2010). How useful are your comments? Analyzing and predicting YouTube comments and comment ratings. In Proc. WWW 2010, pp. 891–900.
9. Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube

recommendations. In Proc. ACM RecSys 2016, pp. 191–198.
10. Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Computer, 42(8), 30–37.
11. Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction, 12(4), 331–370.
12. Zhang, Y., Lai, G., Zhang, M., Zhang, Y., Liu, Y., & Ma, S. (2014). Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In Proc. SIGIR 2014, pp. 83–92.
13. Szabo, G., & Huberman, B. A. (2010). Predicting the popularity of online content. Communications of the ACM, 53(8), 80–88.
14. Ferrara, E., Varol, O., Menczer, F., & Flammini, A. (2013). Traveling trends: social butterflies or frequent fliers? In Proc. COSN 2013, pp. 37–44.
15. Trattner, C., & Jannach, D. (2019). Learning to recommend similar items from human judgments. In Proc. ACM RecSys 2019, pp. 286–294.
16. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1–21.