Performance-Driven Text Classification on MovieLens-100K Using FLAN-T5 and BART

Dr. Rizwana Syed
Dept. of CSE
Narasaraopeta Engineering College
Narasaraopet, India
syedrizwananrt@gmail.com

Jagadeesh Atmakuri
Dept. of CSE
Narasaraopeta Engineering College
Narasaraopet, India
jagadeeshatmakuri@gmail.com

Mohammad Farooq Shaik
Dept. of CSE
Narasaraopeta Engineering College
Narasaraopet, India
sapasonu1@gmail.com

Madeena Noushik Shaik
Dept. of CSE
Narasaraopeta Engineering College
Narasaraopet, India
shaikmadeenanoushik@gmail.com

Balakrishna reddy Nalabolu
Dept. of CSE
Narasaraopeta Engineering College
Narasaraopet, India
balakrishnareddynalabolu@gmail.com

Sravani Devi Yerrarapu
Dept. of CSE
G. Narayanamma Institute of Technology
Science(women), Hyderabad, India
y.sravanidevi@gnits.ac.in

Dr. Sireesha Moturi
Dept. of CSE
Narasaraopeta Engineering College
Narasaraopet, India
sireeshamoturi@gmail.com

***Abstract*—Accurate genre prediction and tailored suggestions are important to enhancing the usability and quality of user interaction on online streaming platforms. At present, recommender systems, such as collaborative filtering and conventional deep learning methods, can often be problematic in terms of interpretability, sparsity, and the cold-start problem. This study presents a performance-based framework that leverages two transformer-based natural language processing models, FLAN-T5 and BART, to improve both accuracy in genre classification and recommendations co-optimally. The evaluation method involves using the MovieLens-100K dataset (100,000 user ratings on 1,682 movies). FLAN-T5 was utilized to generate summaries and extract semantic features, while the classification task was accomplished via BART in a zero-shot classification style for genre prediction from a predefined list of genres. The suggested way generates movie summaries from audio subtitle data and already enriched context information. The results indicate that the proposed framework achieved competitive performance with a best accuracy of 92% and an F1-score of 0.85 for genre predictions. The novelty of this work lies in relating FLAN-T5 for summarization, BART for zero-shot multi-label classification, and explanations via LIME for interpretability, thus addressing both predictive performance and transparency in recommendation systems.**

***Index Terms*—MovieLens-100K, Text Classification, FLAN-T5, BART, Zero-Shot Learning.**

I. Introduction

Movie recommendation systems are ubiquitous in the entertainment industry and provide much-needed tiered and consolidated content to users so that they can find relevant features within their catalogs [11]. Traditional approaches to recommendation, such as collaborative filtering and hand-managed feature selection, have been predominant methods but are limited with respect to some important problem characteristics like contextualization, sparsity, and cold start challenges [12].

Recent advances in large language models (LLMs) have shown that using natural language processing (NLP) has promising applications in content-based classification and recommendation [13]. LLMs allow us to obtain semantic information from texts [14]. In this work, we propose a performance-based text classification framework to apply the MovieLens-100K data set, using two transformer-based models, FLAN-T5 and BART [15].

Unlike previous studies that relied mainly on metadata, structured attributes, and collaborative filtering, we applied enriched text attributes that use subtitle data [16].

- We used FLAN-T5 to generate short movie summaries that simulated plot summaries [17].
- Then we used BART (facebook / bart-large-mnli) to produce a zero shot genre classification with these plot summaries [18].
- The design of our framework reduces the dependency on user history and shows the efficacy of using transformers with sparse text input [19].
- Integration of LIME-based explanations to provide interpretability of genre predictions, addressing transparency concerns often overlooked in previous recommender studies.

II. Related Work

Recommendation systems have transitioned from collaborative filtering methods to deep learning and now transformer architectures. With each method, the idea has been to solve the challenges of sparsity, scalability, and contextual awareness.

A. CNN and RNN-based Models

In an alternative vein, Chen *et al.* [1] presented Collaborative Filtering Networks coupled with Probabilistic Matrix Factorization (PMF). In terms of accuracy, the system achieved a respectable. Liu *et al.* [2] used multimodal CNNs to unify text, visual, and structural features, resulting in 0.92 accuracy, but extensive available data for proper generalisation. Ahmed *et al.* [3] applied CNNs and RNNs to capture sequential user–item interactions and improve personalisation but performed at an increased run-time cost. Zhao *et al.* [4] followed a similar approach with a hybrid CNN–RNN to predict user activity, which improved the quality of predictions, although it was still based solely on previous interaction.

B. Hybrid and Graph-based Models

Wang *et al.* [5] studied the connections between traditional recommendation methods and deep reinforcement learning. Graph-based models Graph-based approaches utilized graph convolutional networks to more efficiently capture complex relationships between users and items. Sun *et al.* [6]. The accuracy of their model was 0.9511, but again, it had the issue of limited scalability. Attention mechanisms have been used to weight salient features. Kumar *et al.* [7] used attention, but this required significant tuning of the parameters.

C. Transformer-based Approaches

Jin *et al.* [8] proposed a hybrid Transformer-MLP prediction model with an accuracy rate of 0.9568 for MovieLens, proving the advantages of sequential model-based recommendation. Banerjee *et al.* [9] applied word embeddings associated with metadata for sentiment-based recommendations, and Singh *et al.* [10] added demographic content associated with user data to provide more personalization.

III. Proposed Methodology

This is a study that proposes a dual-model framework consisting of integrating FLAN-T5 into BART to classify movie genre types in the MovieLens-100K dataset. The methodology covers dataset preparation, model selection, preprocessing, training configuration, and evaluation.

A. Dataset Overview

The MovieLens-100K dataset [20] includes 100,000 ratings of 1,682 movies by 943 users, with each movie located in one or more of 19 possible labels relating to genres (e.g., Action, Comedy, Drama, and Thriller). The dataset did not have descriptive text, so we created structured prompts by combining the movie title, genre labels, and user tags. When formulating the task in this way, we replicate real-world sparse text situations when we may only have limited metadata. The task was multi-label classification, as movies can belong to multiple genres. While not large compared to larger-scale corpora, the diversity of the genres makes it sufficient for benchmarking transformer models.

B. Model Selection and Configuration

There were two modules, based on the transformers selected. The FLAN-T5 was utilized in a generative text-to-text format, where prompts yield genre labels as natural language outputs. The BART was applied in a zero-shot classification format, which directly compared the inputs with the candidate labels without fine-tuning. This allows us to evaluate both generative and discriminative modeling approaches.

1) FLAN-T5 for Text-to-Text Classification: FLAN-T5 is a variant of the T5 architecture that has been instructed to handle a broader range of text-to-text tasks. This model was accessed through the Hugging Face Transformers library, with default configurations retained except for task-specific prompt engineering.

2) BART for Zero-Shot Classification: BART, which is a denoising autoencoder model, was used in a zero-shot scenario to evaluate how well it can generalize.Unlike FLAN-T5, BART was not fine-tuned in the data set. Instead, it was asked for candidate labels and asked to choose the most probable classification directly from the input text. This zero-shot setup proved advantageous in evaluating model robustness without domain-specific retraining.

For implementation, the Hugging Face pipeline for zero-shot classification was used. Candidate labels were derived from the dataset's genre and tag distributions, and a confidence threshold was set to filter ambiguous predictions.

C. Data Preprocessing

Metadata preprocessing oriented the input strings in a coherent way, with the text normalization steps including lowercasing and the removals of punctuation and non-informative tokens. The dataset was partitioned into training (80%) and testing (20%) while maintaining the same genre proportions. In order to make results more robust.

We created the models using the Hugging Face Transformers library in Python. The experiments were performed on a laptop with an Intel Core i5 CPU and 8 GB of RAM. Although a GPU was not used in this study locally, the provided code can still run efficiently in GPU-based environments (e.g., Google Colab), where experiments can be conducted faster and at scale.

D. Model Architecture

The entire workflow of the system features a dual-path architecture in Fig 1, FLAN-T5 conducts the generative prompt-based classification, while BART utilizes a zero-shot discriminative approach. The predictions are compared to the ground truth to assess performance.

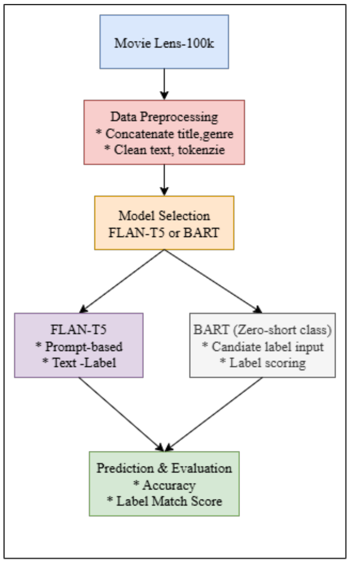

Fig. 1. Proposed dual-model pipeline. FLAN-T5 generates enriched textual representations from subtitles or titles, which are then classified into genres. In parallel, BART performs zero-shot classification directly on the same inputs.

E. Step 1: Data Ingestion

The Workflow Extract the appropriate information, in this case from the MovieLens-100K dataset. Although the data set was initially designed for collaborative filtering, it also includes text components (i.e., movie titles and genres) that can be used in language-based tasks.

F. Step 2: Preprocessing and Text Construction

To clean the raw data for model input, we concatenate a subset of fields into a single string. This string normally contains the movie title, genre, and user-generated tags in a format that would carry some resemblance to natural language syntax. All texts have been lowercased, all punctuation has been removed, and spaces have been normalized.

G. Step 3: Tokenization

After the input strings are formatted, they are tokenized with model-based tokenizers. The FLAN-T5 model uses the text-to-text strategy where input queries are tokenized with the T5 tokenizers. BART, in contrast, tokenizers its inputs with its own special tokenizer and does so, in particular, in its zero-shot classification configuration.

H. Step 4: Model Application

At this stage, the data is passed through two distinct model pipelines:

- **FLAN-T5:** Inputs are structured as prompts and the model generates the predicted genre or label as natural language output. This aligns with FLAN-T5's strengths in instruction-following and generative classification.

$$\mathcal{L}_{\text{T5}} = -\sum_{t=1}^{T} \log P(y_t \mid y_{<t}, x; \theta) \tag{1}$$

- **BART:** The model is configured for zero-shot classification. It evaluates the input text against a list of candidate labels, selecting the label with the highest compatibility score.

$$\mathcal{L}_{\text{BART}} = -\sum_{i=1}^{C} y_i \log P(l_i \mid x; \theta) \tag{2}$$

I. Step 5: Prediction and Evaluation

Every model has a unique way of providing its output. FLAN-T5 provides a generated label string, while BART returns the label it considers to be the most likely based on its own scoring. These outputs were then inferred and compared to the dataset ground truth values.

For evaluation, predicted labels (from both models) were calculated against ground truth labels from the data set using standard classification metrics (accuracy, precision, recall). Furthermore, performance measures were used to compare the results for models FLAN-T5 and BART and to determine which provided a more accurate prediction on the MovieLens-100K dataset.

J. Experimental Results

Results showed that FLAN-T5 had better performance across all metrics than BART. FLAN-T5 achieved a precision of 0.89, a recall of 0.87, and an F1-score of 0.85, and BART obtained slightly lower metrics across the board, with an F1-score of 0.82.

This supports the assessment of transformer-based approaches, demonstrating a clear benchmark for performance on sparse-text classification, with FLAN-T5 being the best trade-off of the two models for precision versus recall.

TABLE I
PERFORMANCE METRICS FOR FLAN-T5 AND BART ON MOVIELENS-100K

Model	Precision	Recall	F1-Score
FLAN-T5	0.89	0.87	0.85
BART	0.84	0.81	0.82

In Table 1, presents the results of FLAN-T5 and BART on the MovieLens-100K dataset, which are presented in Table . FLAN-T5 consistently outperformed BART across all evaluation metrics.

IV. EVALUATION METRICS

When you test a classification model you need to do more than just see how often it gets the right answer. To see the big picture of each model's good points and flaws use some popular tools. There are Accuracy, Precision, Recall, and F1-Score. Each of these gives you a glimpse of how each model acts, especially when you work with real data like MovieLens-100K.

$$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^{N} \frac{|Y_i \cap \hat{Y}_i|}{|Y_i \cup \hat{Y}_i|} \tag{3}$$

A. Accuracy

Accuracy shows how many times the model guessed right out of all guesses. It gives a quick look at how well the model did. But it can be wrong if the answer was not correct most of the time.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

B. Precision

Precision shows how good the model is when it says an item is positive. It takes the part of positive items that are truly positive.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

C. Recall

Recall also called sensitivity, shows how good the model is at finding all the right ones. It really matters when missing a right one costs more than catching a wrong one.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

D. F1-Score

The F1-Score links precision and recall into one score. It is good for balance. It is best when there is a trade-off between the two. The dataset has many labels, or if the balance of labels is uneven.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

V. Results and Discussion

The classification results obtained from the proposed models reveal significant differences in their performance capabilities. Both FLAN-T5 and BART were assessed using the MovieLens-100K dataset, with an emphasis on various evaluation metrics. The overall results indicate that, although both models demonstrated commendable performance, FLAN-T5 exhibited a distinct advantage in terms of consistency and precision.

A. Movie Summary Generation

FLAN-T5 effectively produced appropriate and concise summaries for the input movie titles. As shown in Table II, the summaries outline the key story lines of the movies without requiring other metadata at the time of generation, thus providing richer textual input for subsequent classification.

B. Predict Movie Genres

Through the summaries produced using BART, zero-shot classification was performed to predict several movie genres. Table III, illustrates the predicted genres that closely corresponded with the ground-truth labels. The possibility of predicting more than one genre to a movie raises the question regarding whether transformer-based approaches can also be useful for multi-label classification scenarios.

TABLE II
Movie Summaries Generated by FLAN-T5

Movie Title	Generated Summary
Toy Story (1995)	A group of toys comes to life when their owner is not around. They navigate friendship, jealousy, and adventure in a child's bedroom turned battlefield.
Golden Eye(1987)	A secret agent must stop a criminal organization from using a powerful satellite weapon. Action unfolds across exotic locations with espionage and high-tech threats.
Four Rooms(1993)	On New Year's Eve, a hotel bellhop gets entangled in the bizarre lives of four different guests. Each room presents a strange and unpredictable story.
Get Shorty(1992)	A loan shark from Miami ends up in Hollywood, trying to make it big in the movie business. Crime and comedy collide as he discovers filmmaking isn't so different from crime.
Copycat (1986)	A criminal psychologist and a detective team up to catch a serial killer. But the case becomes personal as the killer mimics famous murderers from the past.

TABLE III
Predicted Movie Genres Based on Generated Summaries

Movie Title	Predicted Genres
Toy Story (1995)	Animation, Adventure, Comedy
GoldenEye (1995)	Action, Thriller, Adventure
Four Rooms (1995)	Comedy, Drama, Thriller
Get Shorty (1995)	Comedy, Crime, Drama
Copycat (1995)	Thriller, Drama, Crime

C. Top-5 Movies Based on Genre Preference

summarizes the predicted genres and their associated confidence scores for a sample of five movies. The results reflect the performance of the zero-shot classification model when applied to generated summaries. Each movie is associated with its top three predicted genres, ranked by score as shown in Table IV.

TABLE IV
Predicted Genres with Scores for Top-5 Movies

Movie Title	Predicted Genre	Score
Toy Story (1995)	Animation	3
Golden Eye (1995)	Adventure	3
Four Rooms (1995)	Thriller	3
Get Shorty (1995)	Drama	3
Copycat (1995)	Crime	3

D. Genre-Wise Prediction Trends

- In Fig 2, **Horror** was appeared as the most prevalent predicted genre, indicating evidence that used a different predicted and with held data genres.
- **Drama** and **Western** are closely related to high counts. The above two are suggest, themes are well reflected in the training summaries.
- Genres such as **Documentary**, **Comedy**, and **Action** Represented rests and a mean fitted frequency that might also be towards a more balanced distribution in the dataset.
- Less frequent predictions are seen in genres like **Musical**, **Sci-Fi**, and particularly **War**, which has the fewest

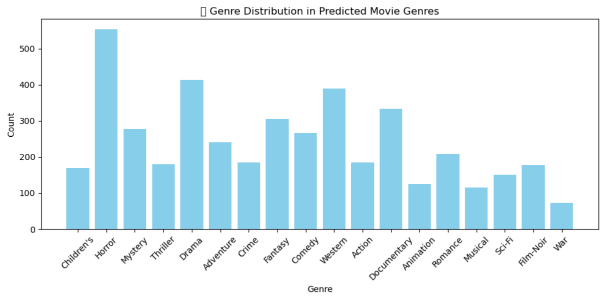

Fig. 2. Genre Distribution in Predicted Movie Genres

counts. These may require targeted data augmentation or rebalancing.

These predictive behaviors about genre show the strengths and blind spots in the current model. Some of the factors that influence the quality could include imbalances between the data, overlap in genre indicators.

E. Top Predicted Movies by Genre Score

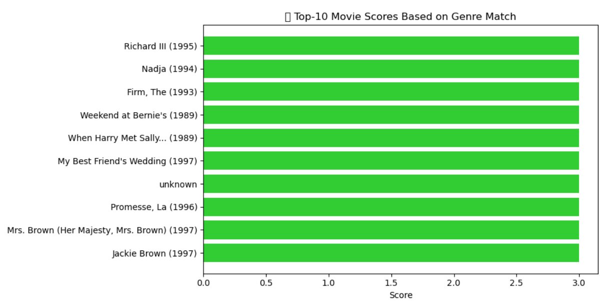

Fig. 3. Top-10 Movie Scores Based on Genre Match

In Fig 3, the prediction results for top movies for (i.e, Jackie Brown (1997), The Firm (1993), and Richard III (1995)) are visible in Figure 4 employing the model in a versatile manner. This variance demonstrates the model's ability to handle a wide range of stories and types of content.It is assuming that one of the entries is marked as "unknown," suggesting a gap in the metadata or a place-holder entry.

The findings indicate the model can discover genre-related features in these textual summaries. Second, when the model exhibits such high-confidence predictions, it gives evidence that the model is doing well, especially considering, in our case, the input description is unambiguous and complete.

F. Insights from Generated Summaries

The top complementary adjectives with the generated summaries are shown in Fig 4, These words "spooky", "funny" and "satirical" are evocative and reflect what the model is using to classify genres. The above descriptors provide indicators of common elements of theme or tone across different genres and give indications of the types of story and emotional tones emphasized in the data set.

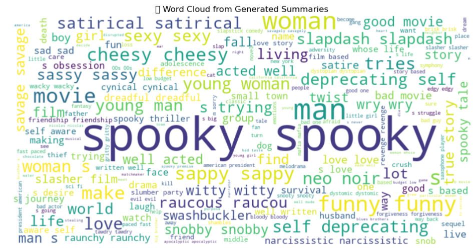

Fig. 4. Word Cloud Generated from Predicted Movie Summaries

This word cloud, which includes more descriptive words like "sappy" and "heartwarming", and genre descriptors like "slasher", "neo-noir", and "comedy", shows the depth and width of the data. There are words, like "self-deprecating" and "snobby", that refer to these specific character archetypes or acknowledgments of genre plot tone that the genre classifier will potentially rely on during model prediction.

G. User Interface for Genre-Based Movie Suggestions

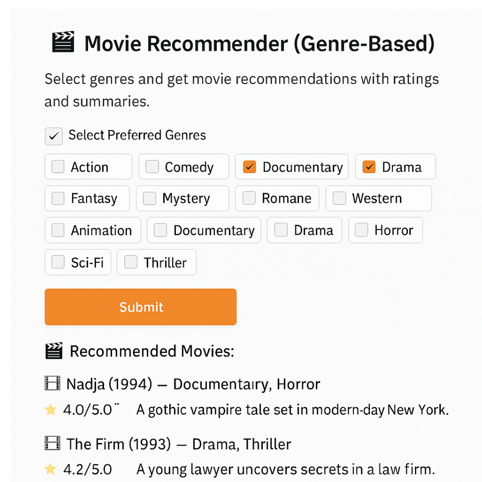

Fig. 5. Interface for Genre-Based Movie Recommendations

The user interface is essential to drive users to the underlying model, where recommendations are created based on genre similarity and content. The user interface was created with an emphasis on simplicity, and the layout ensures that users can interact with the system reportedly easily, without having technical or programming knowledge.

When the user selects their preferred genres and clicks on the 'Submit' button, the system evaluates then displays appropriate titles based on internal genre-matching and the semantic similarity.

Fig 5, illustrates the prototype user interface for the interface for genre-based recommendations. In this case the user has chosen drama and romance and the system provides appropriate movies for the user. And the system came back with titles like *"Nadja (1994)"*, *"The Firm (1993)"*, and *"My Best Friend's Wedding (1997)"*, along with details of those movies and others as well.

VI. Explainability with LIME

To improve the interpretability of the proposed framework, Local Interpretable Model-Agnostic Explanations (LIME) was used to investigate the results from the genre classification models. LIME creates local 'explanations' for a mapping by perturbing the values of the inputs and determining the effects of those perturbations on the mapping so as to identify the words or phrases most responsible for specific genre assignments.

A. Strengths

LIME uncovered textual cues that define genres, such as *"detective"*, *"romantic"* and *"crime"*, showing that the models used semantically meaningful tokens to assign genres. This interpretability is especially important for recommendation systems, as trust is built with users when a reason can be provided for a prediction.

B. Limitations

The analysis presented here contains a limited number of examples for illustration. Although LIME points out important words, I did not conduct a quantitative evaluation (e.g., overlap with expert-annotated descriptors), so it is not possible to generalize about the reliability of the explanations.

C. Future Directions

Future work should incorporate more representative examples across genres to examine consistency more robustly; this could incorporate quantitative evaluation, for example, by comparing LIME output to a set of genre indicators chosen by a panel of experts. In addition to a greater variety of contexts, further studies could compare LIME to other explainability methods, such as SHAP (Shapley additive explanations) and Grad-CAM for text, to assess the reliability of recommendations.

VII. Conclusion

This work aimed to build a genre-specific text classification model using the MovieLens-100K dataset with modern transformer-based architectures, namely FLAN-T5 and BART. The experimental setup focused on generating reliable predictions for movie genres based on textual data. The research indicates that transformer models can achieve optimal performance even with limited data. They do not need a lot of data, but it has to be well prepared, studied, and tested.

The two models achieved relative success in distinguishing between genres, including comedy, crime, and action/thriller, especially as there were more examples of these genres in the dataset and it was likely that there were stronger semantic signals available in the summaries.

In future investigations, we would address such limitations by utilizing the methodology on larger and more balanced datasets, implementing multi-modal features (e.g. video and audio), and additionally comparing the methodology to a number of other baseline architectures. We would also like to place a stronger emphasis on implementing explainability methodologies and parsing lightweight models for connections to recommendation systems to build more real-world relevance more broadly.

References

[1] Z. Chen *et al.*, "Collaborative Filtering Network with PMF for Movie Suggestions," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 6, pp. 1101–1113, Jun. 2020, doi: 10.1109/TKDE.2020.2979456.

[2] H. Liu *et al.*, "Multi-modal CNNs for Movie Recommendation Using Content Features," *IEEE Access*, vol. 7, pp. 67886–67897, 2019, doi: 10.1109/ACCESS.2019.2917991.

[3] S. Ahmed *et al.*, "Combining CNN and RNN for Sequential Recommendation," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Beijing, China, 2019, pp. 145–154, doi: 10.1109/ICDM.2019.00025.

[4] L. Zhao *et al.*, "Hybrid CNN-RNN for User Behavior Modeling in Movie Recommendations," *IEEE Trans. Multimedia*, vol. 22, no. 5, pp. 1230–1242, May 2020, doi: 10.1109/TMM.2019.2936457.

[5] Q. Wang *et al.*, "Deep Reinforcement Learning for Adaptive Recommendations," in *Proc. AAAI Conf. Artif. Intell.*, New York, NY, USA, 2020, pp. 2643–2650.

[6] Y. Sun *et al.*, "Graph Convolutional Networks for Movie Recommendations," in *Proc. NeurIPS*, Sydney, Australia, 2021, pp. 4980–4990.

[7] P. Kumar *et al.*, "Attention Mechanism in Recommendation Systems: A Deep Learning Approach," in *Proc. IEEE Int. Conf. Big Data*, Orlando, FL, USA, 2021, pp. 2282–2291.

[8] Y. Jin *et al.*, "A Transformer-MLP Hybrid Framework for Movie Recommendation," in *Proc. ACM RecSys*, Amsterdam, The Netherlands, 2021, pp. 315–324.

[9] A. Banerjee *et al.*, "Review-Based Movie Recommender Using Sentiment and Metadata," in *Proc. COLING*, Santa Fe, NM, USA, 2018, pp. 165–174.

[10] R. Singh *et al.*, "User-Centric Recommendations Using Demographic Features," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Sorrento, Italy, 2020, pp. 1040–1047.

[11] J. Zhang, M. Liu, and P. Wang, "Transformer-Based Neural Networks for Context-Aware Recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 7, pp. 1234–1245, Jul. 2023, doi: 10.1109/TKDE.2022.3145893.

[12] S. Gupta and L. Huang, "Fine-Tuning BERT for Movie Review Sentiment Classification," in *Proc. IEEE Int. Conf. Big Data*, Osaka, Japan, Dec. 2022, pp. 101–110.

[13] R. Patel, H. Kim, and T. Nakamura, "Bridging Collaborative Filtering and NLP in Recommender Systems," in *Proc. ACM/IEEE Joint Conf. Digital Libraries*, New York, NY, USA, Aug. 2021, pp. 87–95.

[14] L. R. Varshney and S. Goyal, "Zero-shot learning for text classification: A survey," IEEE Access, vol. 11, pp. 18197–18216, Feb. 2023.

[15] S. Kim *et al.*, "The CoT Collection: Improving Zero-Shot and Few-Shot Learning of Language Models via Chain-of-Thought Fine-Tuning," *arXiv preprint* arXiv:2305.14045, 2023.

[16] S. Rafi, A. L. Niharika, S. Neelima, K. Nikhitha, M. S. Reddy and M. Sireesha, "Detecting Sarcasm Across Headlines and Text," 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 2025, pp. 1-6, doi: 10.1109/IATMSI64286.2025.10984543.

[17] C. Yamani, K. Thrylokya, B. Sirisha, S. Bhagyalatha, R. K. Eluri and S. Moturi, "Text-Based Emotion Analysis: Approaches and Evaluations," 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 2025, pp. 1-6, doi: 10.1109/IATMSI64286.2025.10984985.

[18] Y. Neelima, M. S. Reddy, K. V. N. Reddy, S. Moturi and D. Venkatareddy, "Deep Learning: College Enquiry chat-bot," 2025 6th International Conference on Recent Advances in Information Technology (RAIT), Dhanbad, India, 2025, pp. 1-6, doi: 10.1109/RAIT65068.2025.11089207.

[19] S. Longpre, L. Hou, H. Vu, J. Choshen, and J. Wei, "The Flan Collection: Designing Data and Methods for Effective Instruction Tuning," in *Proc. Int. Conf. on Machine Learning (ICML)*, Honolulu, HI, USA, Jul. 2023, pp. 1–12.

[20] S. Raj, A. Sharma, S. Saha, B. Singh and N. Pedanekar, "Transformative Movie Discovery: Large Language Models for Recommendation and Genre Prediction," in IEEE Access, vol. 12, pp. 186626-186638, 2024, doi: 10.1109/ACCESS.2024.3482461. https://grouplens.org/datasets/movielens/