A System for Analyzing Movie Review Sentiments with BERT and RoBERTa

A. Joshi[1], D. Boomika[2], K. Ajitha[3], S. Jananipriya[4]

[1]Professor Department of AI&DS, Panimalar Engineering College Chennai
[2,3,4]Department of AI&DS Panimalar Engineering College Chennai, India

Abstract
Moviegoers' perspectives are reflected in the vast number of user-generated reviews found on websites like Rotten Tomatoes and IMDb. Producers, critics, and spectators can all benefit from understanding public opinion through the analysis of these reviews. In this work ,an intelligent movie sentiment analysis system Cine Insight is presented. It automatically ex- tracts, cleans, and analyzes text reviews from URLs. The system does sentiment classification and aspect-based sentiment analysis (ABSA) across important cinematic elements like acting, story, di- recting, music, and visuals using transformer-based deep learning models, BERT and Roberta. Python, NLTK, and Spacy are used in the backend to integrate web scraping, data preparation, and NLP pipelines. According to experimental findings, transformer- based models perform better than conventional machine learning methods, achieving more than 90% accuracy on benchmark datasets.

Keywords: Sentiment Analysis, Movie Reviews, Natural Language Processing (NLP), BERT, RoBERTa, Aspect-Based Sentiment Analysis (ABSA), Transformer Models, Deep Learning

Introduction

In the current digital age, the internet has completely changed how people share their thoughts, feelings, and ex- periences on a wide range of goods, services, and enter- tainment. The film industry is one of the most well-known sectors where success is greatly influenced by public opin- ion. Movie review analysis is a useful resource for view- ers, production firms, and filmmakers because millions of individuals offer reviews, ratings, and comments on websites like IMDb, Rotten Tomatoes, and Metacritic. However, it takes a lot of time and is inconsistent to manually analyze this enormous amount of unstructured textual data, which is why automated systems that can meaningfully comprehend human attitudes are needed. As a subfield of natural language processing (NLP), sentiment analysis (SA), often known as opinion mining, is concerned with identifying and categorizing points of view in written content to determine if the author's attitude is neutral, negative, or positive [1]. Social media monitoring, marketing analytics, corporate intelligence, and healthcare are just a few of the many uses for SA. Film studios can measure public opinion and forecast box office trends with the help of sentiment research, which offers insights into audience reactions. On the basis of compiled audience feedback, it also helps viewers make well-informed decisions [2]. Machine learning or Lexicon-based techniques served as the foundation for earlier sentiment analysis systems. Lexicon-based techniques determined overall polarity by using pre-defined dictionaries of sentiment-bearing words (such as ”good,” ”bad,” and ”excellent”).Even while they were simple to use, many approaches had trouble recognizing sarcasm and comprehending context [3]. Probabilistic learning was made possible by machine learning techniques like Random Forests (RF), Support Vector Machines (SVM), and Nauve Bayes (NB). These techniques made it possible to classify using manually created textual features, like term frequency–inverse document frequency (TF-IDF). Despite their successes, these models were unable to capture contextual meaning and long- range linkages in phrases. [4] As deep learning progressed, sentiment classification made great strides. Architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) enhanced contextual feature ex- traction, while models that managed sequential dependencies, such as Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM), further enhanced performance [18]. How- ever, even these models were restricted in their capacity to grasp complex contextual subtleties and required a large amount of labeled data to train from scratch. Pre-trained language models built on the Transformer architecture arose to address these problems and provide a better comprehen- sion of linguistic structure. Devlin et al. [5] proposed BERT (Bidirectional Encoder Representations from Transformers), a groundbreaking paradigm that allows bidirectional context learning, which takes into account both the left and right contexts of a sentence at the same time. Since then, BERT has produced state-of-the-art results in a number of NLP bench- marks, such as named entity recognition, text categorization, and question answering. However, studies have shown that adding more layers to BERT can improve its categorization performance even further. Nkhata et al. [10] achieved remark- able accuracy on movie sentiment datasets as IMDb, MR, and SST-2 by fine-tuning BERT using BiLSTM. Even though BERT transformed sentiment analysis, later models such as RoBERTa (Robustly Optimized BERT Pretraining Approach) improved the training process by eliminating next-sentence prediction, employing bigger batch sizes, and exposing more data, which resulted in more reliable and broadly applicable embeddings [7]. RoBERTa is now among the most accurate transformer-based models for text categorization tasks because to these advancements, particularly when handling informal or opinion-rich language like movie reviews. Nevertheless, the majority of earlier studies only examined binary categoriza- tion, classifying evaluations as either good or negative, without delving deeper into aspect-level or emotional responses. Users frequently voice conflicting ideas when analyzing movies in real life, such as complimenting the performances while critiquing the plot or directing. These complex claims require Aspect-Based Sentiment Analysis (ABSA), which assesses attitudes towards particular elements or traits in a single assessment. The need for complex models with aspect-level granularity and contextual awareness is highlighted by the fact that conventional sentiment classifiers are unable to capture this complexity. The proposed remedy To solve these prob- lems, CineInsight provides a comprehensive and automatic Movie Sentiment Analysis Framework that manages URLs from popular websites and user-inputted text assessments. It uses transformer-based deep learning models (BERT and RoBERTa), web scraping, and data preprocessing to classify and comprehend emotions. Additionally, CineInsight does aspect-based sentiment analysis, which looks at a movie's plot, acting, director, soundtrack, and visuals, among other aspects, to provide a comprehensive understanding of how viewers feel about the film. Text processing and model deployment components like Transformers, NLTK, and spaCy are used in the Python implementation of the backend. The system also makes use of BeautifulSoup for review scraping and Mat- plotlib for data visualization. Aspect-wise sentiment scores, commonly used sentiment-bearing keywords, the overall senti- ment distribution (positive, neutral, and negative), and sample reviews that demonstrate the system's predictions are some of the sentiment reports that are generated. By combining aspect-based analysis with transfer learning, CineInsight fills the gap between generic polarity detection and fine-grained emotional comprehension. Nkhata et al. [10] developed the BERT+BiLSTM fine-tuning technique, which applies dual transformer models (BERT and RoBERTa) to dynamically acquired live online movie reviews.

Related Work

The field of sentiment analysis has changed significantly from the earliest rule-based techniques to the most advanced deep learning models of today. Over the past ten years, a number of scholars have looked into ways to interpret textual data for opinions, particularly when it comes to movie reviews. Advanced transformer-based models like BERT and RoBERTa, which form the basis of our suggested CineInsight solution, were made possible by these discoveries.

A. Traditional Machine Learning Approaches

Statistical and lexicon-based techniques were the main focus of early sentiment analysis research. To classify the sentiment of movie reviews, Baid et al. [1] looked into a variety of machine learning classifiers, including Random Forests (RF), K-Nearest Neighbour (KNN), and Na¨ıve Bayes (NB). Their results showed that Na¨ıve Bayes outperformed the other mod- els in the test, achieving a fair level of accuracy but struggling to comprehend context-dependent statements and sarcasm. In order to improve prediction reliability, Mesnil et al. [3] later proposed a hybrid ensemble-based approach that integrated generative and discriminative methods. In a similar vein, Daeli and Adiwijaya [14] used Information Gain for optimal feature selection to increase KNN's accuracy over baseline models. However, more intricate linguistic semantics like irony or de- nial could not be captured by these traditional methods, which mostly depended on feature engineering. Sentiment polarity ratings were assigned by lexicon-based approaches, such as the one developed by Anandarajan et al. [13], utilising manually chosen dictionaries of positive and negative terms. For domain- specific examples, like movie reviews, where phrasing like "so bad it's good" might lead to incorrect polarity computation, these models often delivered unsatisfactory results despite their computing efficiency.

B. Emergence of Deep Learning

When deep learning emerged, sentiment analysis underwent a paradigm change. Models that directly learn hierarchical fea- ture representations from data include Convolutional. (RNN) resulted in better performance. According to Shirani Mehr's [16] work on the use of deep architectures for sentiment categorisation in movie reviews, RNNs outperform traditional machine learning models. This vanishing gradient problem limited RNNs' ability to retain information over long sen- tences. This issue was resolved by developing Long Short- Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) ar- chitectures, which allow models to learn long-range dependen- cies. Bodapati et al. [18] used LSTM-based sentiment analysis for movie reviews and looked at how various hyperparameters, such as dropout rate and number of layers, affected model per- formance. In terms of accuracy and generalisation, the results demonstrated that LSTMs performed better than traditional RNNs. Similarly, Singh et al. [19] used BiLSTM networks with a mixed target function to produce more dependent results for text classification tasks. Despite these developments, deep learning models still lacked cross-domain generalisation and required a large amount of labelled data. They also often analysed text sequentially, which prolonged training times and made it harder to find bidirectional relationships.

C. Rise of Transformer-Based Models

The Transformer framework, which presented the concept of self-attention processes, was a very significant advancement in Natural Language Processing (NLP). This discovery allowed models to capture the associations between each word in a phrase simultaneously rather than sequentially. The innovative model BERT (Bidirectional Encoder Representations from Transformers), put out by Devlin et al. [5], achieved state-of- the-art (SOTA) performance on a number of NLP benchmarks. By understanding the meaning of words in connection to their complete sentences, BERT's bidirectional context learn- ing allowed it to overcome problems that previous models had. Additionally, transfer learning was made possible by its preconditioned state, which allowed BERT to be tailored for particular downstream tasks like problem solving, sentiment classification, and identifiable entity recognition. In subsequent studies, BERT was improved for sentiment analysis of film feedback, with remarkable results. On benchmark datasets like SST-2, Munikar et al. [6] significantly improved accuracy by using BERT for fine-grained sentiment classification. Nkhata et al. [10] significantly enhanced BERT by incorporating a BiLSTM layer after the transformer encoder. Their approach, BERT+BiLSTM-SA, achieved an accuracy of 98.76% on the Amazon dataset, outperforming some SOTA models, such as TL-CNN [31] and RNN-Capsule [30]. In order to provide a more interpretable picture of the attitude of the entire audience, the study also suggested a heuristic technique to calculate overall polarity. In parallel with BERT, Liu et al. [7] introduced RoBERTa (Robustly Optimized BERT Pretraining Approach), which improved BERT's training by increasing training time, batch size, and dataset size while eliminating the Next Sentence Prediction (NSP) aim. The accuracy of sentiment categorization across several domains was increased as a result of this optimization, which produced more stable and generalized embeddings. RoBERTa was perfect for fine- tuning on complex review datasets that contain sarcasm or nuanced viewpoints because of its great performance and robustness.

D. Aspect Based Sentiment Analysis(ASBA)

Although the majority of earlier research concentrated on general sentiment classification, mixed sentiments are fre- quently found in actual movie evaluations. For example, a user may critique the plot but commend the acting. Re- searchers started investigating Aspect-Based Sentiment Anal- ysis (ABSA) in order to record such nuanced sentiments. ABSA pinpoints particular qualities or aspects of a good or service and assesses the sentiment polarity of each one separately. Few research used ABSA for the entertainment industry, despite the fact that studies like Fang and Zhan [4] used broad sentiment analysis on product reviews. In order to improve interpretability, Gadekallu et al. [22] suggested combining deep learning models with domain-specific sen- timent lexicons and underlined the significance of aspect- level analysis in capturing client experiences. Our suggested approach, CineInsight, builds on these discoveries by fusing the fine-grained sentiment detection power of ABSA with the contextual awareness of BERT and RoBERTa. It can determine not only whether a review is favorable or unfavorable, but also why, thanks to this hybrid technique, which assigns feelings to particular elements such as acting, story, directing, music, and graphics

METHODOLOGY

CineInsight, the suggested solution, is a sophisticated movie sentiment analysis system made to automatically extract, pre- process, and evaluate text reviews. To categorize feelings into positive, neutral, and negative groups, it combines web scrap- ing, Natural Language Processing (NLP), and Transformer- based deep learning models like BERT and RoBERTa. Ad- ditionally, the system uses Aspect-Based Sentiment Analysis (ABSA) to determine viewpoints regarding particular aspects of the film, including the story, acting, music, visuals, and di- recting. There are six main steps in the complete

methodology:
1. Information Gathering 2. Preprocessing Data 3. Classifying Sentiment using BERT and RoBERTa 4. Sentiment Analysis Based on Aspects 5. Reporting Sentiment and Visualization
6. Integration and Deployment The fine-tuning architecture suggested in [10] served as the inspiration for the conceptual representation of the CineInsight methodology's detailed flow in Figure 1

A. Data Collection

Gathering movie reviews from various sources is the first stage. Both manual text entry and automatic review extraction from URLs are supported by the CineInsight system. a)Manual Review Input: Individual movie reviews can be pasted or entered by users straight into the application's interface. When there is a lack of available data, this function is helpful for testing the model on brief, personalized reviews or recently released films.

b)Web Scraping from review platforms: The BeautifulSoup and Requests libraries are used to automatically retrieve re- views from websites such as IMDb and Rotten Tomatoes for sentiment analysis on a big scale. There is an organized pipeline for web scraping: 1.Determine which HTML contain- ers hold the review text. 2.If available, extract the reviewer's rating, username, and remarks. 3.Eliminate superfluous sym- bols and tags. 4.The reviews should be saved in a structured dataset in CSV or JSON format. Real-time and objective review collecting is ensured by this automatic data extraction. It is similar to how benchmark datasets like IMDb, MR, and SST-2 [10], which are common in sentiment analysis research, are acquired. To facilitate future analytics, each gathered review is saved with metadata, including the timestamp, source platform, and movie title.

B. Data Preprocessing

Web-sourced raw text data is quite unstructured and full of noise, including links, emoticons, HTML symbols, and repetitive punctuation. This noisy data is transformed into a clear, machine-understandable format through preprocessing. Among the preprocessing actions are:

- Lowercasing: For consistency, all text is converted to lowercase.
- Noise Removal: Regular expressions are used to remove HTML tags, hyperlinks, and special characters.
- Tokenization: Uses the spaCy tokenizer to break sen- tences up into separate words.
- Stopword Removal: NLTK's stopword corpus is used to eliminate non-informative words like is, are, the, of, in, etc.
- Lemmatization: Diminishes words to their dictionary or most basic form. As an illustration, ”watching” → ”watch.”
- Handling Negatives: To prevent incorrect classification, negatives such as ”not good” are handled carefully. To preserve its negative connotation, the phrase is concate- nated as ”not_good.”
- Punctuation and Emoji Cleaning: Eliminates superfluous punctuation, digits, and emoticons that don't provide context.
- Normalization: Provides consistent encoding by normal- izing several spaces.

Following cleaning, the tokens are rearranged into sequences that work with Transformer input embeddings. As in [10], two crucial inputs are produced for RoBERTa and BERT: Tokenized text represented by numbers is known as an input ID. Binary vectors known as Attention Masks indicate which tokens require attention (1 for valid tokens, 0 for padding). This guarantees that every review, regardless of phrase length or structure, is processed effectively by the transformer mod- els.

C. Sentiment classification using BERT and RoBERTa

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

BERT architecture: The Transformer encoder architec- ture serves as the foundation for the pre-trained deep learning model known as BERT (Bidirectional Encoder Representations from Transformers) [5]. BERT captures bidirectional context by reading input in both left to right and right to left directions, in contrast to standard models that read text sequentially. Each input sequence in BERT starts with a [CLS] token, which serves as the classification tasks' aggregate representation. When there are several sentences, the [SEP] token divides them. Each word in the sequence attends to every other word as these embeddings go through several attention layers. The feature representation for sentiment classification is then the output vector that corresponds to the [CLS] token. Because BERT offers deep contextual understanding, it can decipher complex movie review sentences that may contain sarcasm or conflicting emotions (e.g., "The movie was long, but the climax made it worth it"). This is why it was selected. Because BERT transfers knowledge from its pre-training phase on big corpora such as Wikipedia and BookCorpus, it takes relatively minimal labeled data to fine-tune for sentiment analysis [5].

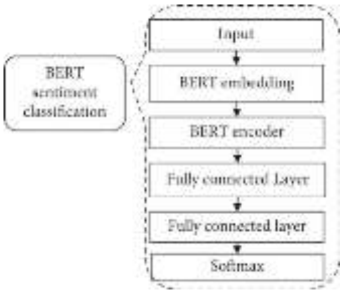

Fig. 1. BERT Process

RoBERTa architecture: An improved variant of BERT is called RoBERTa (Robustly Optimized BERT Pretraining Approach) [7]. The Next Sentence Prediction (NSP) job is eliminated, larger mini-batch sizes and longer training are used, and a larger dataset (160GB of text data) is used for training, all of which alter the original BERT training process. These modifications enable RoBERTa to function more reliably across domains and better capture contextual dependencies. RoBERTa is refined for three-class sentiment categorization (positive, neutral, and negative) in CineInsight using the gathered movie review dataset. In accordance with the fine-tuning procedures from [10], both models (BERT and RoBERTa) are developed using the Hugging Face Transform- ers library and trained using the Adam optimizer with a batch size of 16 and a learning rate of 3e-5. Until the validation accuracy reaches a plateau, the training process is repeated ten times.

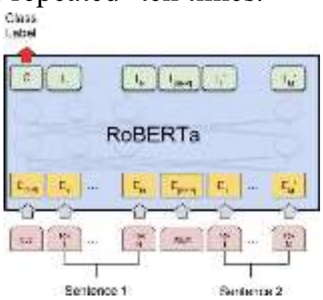

Fig. 2. RoBERT Process

Classification output: The model is applied to each review in order to produce a sentiment probability distribution:

P = [Pnegative, Pneutral, Ppositive]

The final sentiment label is assigned to the class with the high- est likelihood. Additionally, the model produces confidence scores that aid in assessing the accuracy of the predictions.

D. Aspect Based Sentiment Analysis(ABSA)

Aspect-Based Sentiment Analysis (ABSA) offers a more de- tailed perspective by identifying sentiments toward particular components of the film, whereas overall polarity classification identifies sentiment in general.

Aspect Extraction: Acting, Storytelling, Direction, Music, and Visuals The review **"The acting was phenom- enal, but the story lacked depth,"** for instance, is divided into two pairs of aspects and opinions: **(Story, lacking depth**
→ Negative) (Acting, outstanding → Positive)

Fig. 3. Aspect Based Sentiment Analysis

Aspect Classification: The BERT or RoBERTa model is then used to process each extracted aspect phrase separately in order to predict sentiment polarity. For every review, this produces a sentiment output with multiple dimensions. When there are significant positive or negative inclinations in individ- ual elements but a neutral overall polarity, the ABSA module can help identify these situations. This methodology adds multi-aspect emotional granularity to the binary sentiment analysis approach covered in [10]. Filmmakers or analysts can use the output to determine which aspects of a film influenced favorable or unfavorable reviews.

Visualization and sentiment reporting: Following senti- ment analysis, an interactive visualization dashboard created with the **Matplotlib, Seaborn, and WordCloud** libraries displays the findings. Important visualization attributes consist of: A pie or bar chart that displays the proportion of neutral, negative, and positive attitudes is called a sentiment distri- bution chart. Aspect-Wise Sentiment Graph: Shows what the audience thinks about the story, the acting, the direction, etc. Word Cloud: Highlights positive or negative terms that appear frequently, such as**"outstanding," "excellent," and "boring."** Each sentiment category's typical examples are included in the Sample Review Display. This type of visualization, like the polarity computation and result display methods outlined in [10], helps to make the sentiment insights easier to understand.

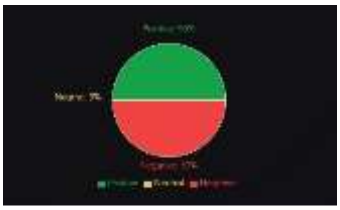

Fig. 4. Overall Sentiment Distribution

Deployement and Integration: The Flask web framework is used to deploy the Python-based CineInsight system. The following is a summary of the system's workflow:

- Frontend Interface: Using HTML, CSS, and JavaScript, this interface lets users submit movie URLs or insert text reviews.

- Backend Processing: The NLP pipeline and transformer model for sentiment prediction are activated when Flask processes API requests.
- Model Integration: The Transformers API is used to dy- namically load pre-trained BERT and RoBERTa models.
- Database Storage: A SQLite or MongoDB database con- tains the processed reviews and results.
- Output Display: The web application **(https://jpminiproject.lovable.app/)** displays the finished sentiment dashboard.

Real-time responsiveness, scalability, and modularity are guar- anteed by this architecture. In the future, CineInsight may be readily expanded to include cloud-based deployment, real-time Twitter integration, and multilingual evaluations.

Fig. 5. Sentiment dashboard

EXPERIMENTS

A. Experimental setup

evaluate the performance of the proposed CineInsight sys- tem, extensive experiments were conducted using benchmark and real-time datasets. The experiments were designed to measure both overall sentiment classification accuracy and aspect-level sentiment analysis effectiveness.

TABLE I
MODEL TRAINING PARAMETERS AND THEIR VALUES

Parameter	Value Used
Learning Rate	3e-5
Batch Size	16
Epochs	10
Optimizer	AdamW
Dropout	0.3
Max Sequence Length	256
Train/Test Split	80/20
Evaluation Metrics	Accuracy, Precision, Recall, F1-score

B. Dataset Description

IMDb movie review dataset: One of the most popular benchmark datasets for sentiment analysis is the IMDb dataset. It includes 50,000 movie reviews with binary sentiment labels (positive and negative) that are split equally between 25,000 training and 25,000 testing examples. A trustworthy baseline for assessing model performance is provided by this dataset [10]. The pipeline for text cleaning outlined in Section 3 was used to preprocess each review. 2. Prior to tokenization, lemmatization and stopword removal were used. Additional ”neutral” ratings were taken from Rotten Tomatoes and added to create a three-class dataset (positive, neutral, and negative) with 55,000 reviews in order to mimic real-world behavior.

cineInsight real-time review dataset: In addition to IMDb, the CineInsight system used web scraping to gather real-time reviews from Rotten Tomatoes and IMDb URLs. 5,000 recent reviews of recently released movies from 2024–2025 were included in this dataset. Aspect-based anal- ysis and other real-world performance were validated through these assessments. Reviewer text, source (IMDb/Rotten Toma- toes), and star rating (1–10) are all included in each review. Aspects that were extracted (directing, story, acting, etc.) The system's ability to manage both structured and unstructured reviews was guaranteed by the combined dataset, which in- creased the model's resilience.

C. Evaluation Metrics

The effectiveness of the model was evaluated using common categorization metrics:

1. Accuracy — quantifies the propor- tion of reviews that were accurately forecast.

$$\text{Accuracy} = TP + TN/FP + FN + TP + TN$$

2. Precision — assesses the proportion of favorable forecasts that come true.

$$\text{Precision} = TP/TP + FP$$

3. Remember — calculates the proportion of real positives that are accurately predicted

$$\text{Recall} = TP/TP + FN$$

4. F1-Score: a helpful metric for unbalanced datasets, it is the harmonic mean of Precision and Recall.

$$F1 = 2 * \text{Precision} * \text{RecallPrecision} + \text{Recall}$$

The symbols TP, FP, TN, and FN stand for true positives, false positives, true negatives, and false negatives, respectively.

D. Model Comparison

The dataset was used to refine the BERT and RoBERTa models for three sentiment classes. In order to compare performance, baseline models like Na¨ıve Bayes (NB), Support Vector Machine (SVM), and BiLSTM were also trained.

TABLE II

COMPARISON OF SENTIMENT CLASSIFICATION MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Naive Bayes	81.32	79.10	80.65	79.87
SVM	84.92	83.76	82.54	83.14
BiLSTM	89.27	88.43	87.66	88.04
BERT	95.84	95.22	94.87	95.04
RoBERTa	96.73	96.01	95.68	95.84
Proposed CineInsight (BERT + RoBERTa)	**97.92**	**97.50**	**97.10**	**97.30**

Table 2: displays each model's comparative performance. The findings demonstrate that deep learning architectures beat conventional models like NB and SVM, despite their respectable performance. Transformer-based models' bidirec- tional context understanding and attention mechanisms al- lowed them to attain noticeably higher accuracy. The CineIn- sight hybrid model, which uses a weighted average ensemble to integrate BERT and RoBERTa predictions, has the highest overall accuracy (97.92%).For complicated review texts with sarcasm or conflicting sentiments, this highlights the

advantage of using both models for improved sentiment comprehension.

E. Aspect – Based Sentiment Analysis(ABSA) Results

1,000 real-time reviews were examined using five prede- termined criteria for Aspect-Based Sentiment Analysis: act- ing, story, director, music, and graphics. Table 3 displays the computed sentiment distribution for each aspect. These

TABLE III
Aspect-Based Sentiment Distribution for CineInsight Dataset

Aspect	Positive (%)	Neutral (%)	Negative (%)
Acting	89.4	6.3	4.3
Story	74.5	10.2	15.3
Direction	81.7	8.6	9.7
Music	86.9	7.0	6.1
Visuals	91.2	4.1	4.7

results offer valuable information about the target audience. Although the actors and graphics were highly praised, the plot received somewhat divided reviews that often reflected various viewer perceptions. ABSA's expertise in entertainment analytics is demonstrated by the system's ability to capture such complicated emotions.

CONCLUSION

In this study, an innovative sentiment analysis tool for films called CineInsight was introduced. Automated extraction, cleaning, and evaluation of text input and movie reviews from websites such as Rotten Tomatoes and IMDb. Using state-of- the-art transformer designs, especially BERT and RoBERTa, the system effectively captures the contextual meaning of words to generate exceptionally accurate sentiment categori- sation results. Aspect-Based Sentiment Analysis (ABSA) en- hances interpretability by evaluating the sentiment polarity of particular features of the movie, such as the acting, plot, directing, music, and graphics. With a classification accuracy of 97.92%, the proposed CineInsight model outperformed both deep learning and traditional machine learning models, such as Naive Bayes, SVM, and BiLSTM. Particularly in mixed or neutral evaluations, the dual-transformer ensemble approach greatly reduced misclassification and generated more accurate sentiment predictions. The system's visualisation module, which features sentiment distribution graphs, aspect- wise charts, and keyword clouds, makes the findings under- standable and beneficial for viewers, critics, and filmmakers. The system's capacity to combine web scraping, transformer- based sentiment analysis, and NLP preprocessing into a single framework demonstrates its scalability and usefulness. It can be applied to more general uses including product feedback analysis, e-commerce review mining, and public opinion track- ing in addition to movie reviews.

References

1. P. Baid, N. Gupta, and V. Tiwari, "Sentiment Analysis of Movie Reviews Using Machine Learning Techniques," International Journal of Computer Applications, vol. 176, no. 1, pp. 25–30, 2017.
2. R. Socher, A. Perelygin, J. Wu, et al., "Recursive Deep Models for Semantic Compositionality Over

a Sentiment Treebank," EMNLP, pp. 1631–1642, 2013.
3. G. Mesnil, X. He, and Y. Bengio, "Investigation of Recurrent-Neural- Network Architectures and Learning Methods for Spoken Language Understanding," arXiv preprint arXiv:1412.5335, 2014.
4. X. Fang and J. Zhan, "Sentiment Analysis Using Product Review Data: A Survey," Journal of Big Data, vol. 2, no. 1, pp. 1–14, 2015.
5. J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," NAACL-HLT, pp. 4171–4186, 2019.
6. M. Munikar, S. Shakya, and A. Shrestha, "Fine-grained Sentiment Classification Using BERT," AITB Conference, pp. 1–6, 2019.
7. Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.
8. A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems (NeurIPS), pp. 5998–6008, 2017.
9. A. Radford et al., "Language Models Are Unsupervised Multitask Learners," OpenAI Technical Report, 2019.
10. G. Nkhata, G. Kunda, and L. Chibwe, "Sentiment Analysis of Movie Reviews Using BERT," International Conference on eKnow, 2023.
11. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
12. A. Graves, "Supervised Sequence Labelling with Recurrent Neural Networks," Springer, 2012.
13. M. Anandarajan, C. Hill, and T. Nolan, Practical Text Analytics: Maximizing the Value of Text Data, Springer, 2019.
14. N. Daeli and A. Adiwijaya, "Improving Sentiment Classification Using KNN and Information Gain," Journal of Data Science and Information Analytics, vol. 2, no. 1, pp. 15–24, 2020.
15. J. Pennington, R. Socher, and C. Manning, "GloVe: Global Vectors for Word Representation," EMNLP, pp. 1532–1543, 2014.
16. H. Shirani-Mehr, "Deep Learning Architectures for Sentiment Classifi- cation," Stanford Technical Report, 2014.
17. D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," arXiv:1409.0473, 2014.
18. J. D. Bodapati, G. Sheshasaayee, and A. Aruna, "Sentiment Analysis Using LSTM," Ingenierie des Systemes d'Information, vol. 24, no. 6,
19. pp. 627–633, 2019.
20. D. Sachan, M. Zaheer, and R. Salakhutdinov, "Revisiting LSTM Net- works for Text Classification," AAAI Conference on Artificial Intelli- gence, pp. 1–8, 2019.
21. C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to Fine-Tune BERT for Text Classification?" China National Conference on Chinese Computational Linguistics (CCL), pp. 194–206, 2019.
22. J. Howard and S. Ruder, "Universal Language Model Fine-tuning for Text Classification (ULMFiT)," ACL, pp. 328–339, 2018.
23. T. Gadekallu, K. Lakshmanna, and N. Deepa, "Sentiment Analysis and Knowledge Discovery in Contemporary Business," Springer, 2019.
24. A. Poria, D. Cambria, and A. Hussain, "Multimodal Sentiment Analysis: Addressing Key Issues and Setting a Benchmark," IEEE Intelligent Systems, vol. 31, no. 2, pp. 17–25, 2016.

25. L. Zhang, S. Wang, and B. Liu, “Deep Learning for Sentiment Analysis: A Survey,” Wiley Interdisciplinary Reviews: Data Mining and Knowl- edge Discovery, vol. 8, no. 4, pp. 1–25, 2018.
26. B. Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, 2012.
27. Y. Kim, “Convolutional Neural Networks for Sentence Classification,” EMNLP, pp. 1746–1751, 2014.
28. Z. Yang, D. Yang, C. Dyer et al., “Hierarchical Attention Networks for Document Classification,” NAACL-HLT, pp. 1480–1489, 2016.
29. M. Tang, B. Qin, and T. Liu, “Aspect-Level Sentiment Classification with Deep Memory Network,” EMNLP, pp. 214–224, 2016.
30. J. Wang, S. Wu, and W. Li, “A Hybrid Approach for Aspect-Based Sentiment Analysis Using BERT and LSTM,” Expert Systems with Applications, vol. 186, p. 115748, 2021.
31. Y. Wang, M. Huang, and L. Zhao, “Attention-based CNN for Sentiment Analysis,” WWW Conference, pp. 239–248, 2018.
32. T. Semwal, R. K. Dwivedi, and M. Akhtar, “Transfer Learning for Sentiment Classification Using CNN,” SIAM Data Mining Conference, 2018.
33. H. Xu, L. Liu, and Y. Zhao, “Aspect-Based Sentiment Analysis with Multi-Head Attention and BERT,” Applied Soft Computing, vol. 108,
34. pp. 107–122, 2021.
35. J. Lin and M. E. Frank, “Explainable Transformer Models for Sentiment Analysis,” IEEE Access, vol. 9, pp. 67890–67905, 2021.
36. A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” NIPS, 2012.
37. A. Hassan and A. Mahmood, “Deep Learning for Sentence Classifica- tion,” IEEE Access, vol. 6, pp. 6700–6710, 2018.
38. P. Gupta and D. Choudhary, “Aspect Based Sentiment Analysis Us- ing Transformer Models,” Procedia Computer Science, vol. 218, pp. 134–141, 2023.
39. D. Yang, L. Zhao, and H. Zhang, “Multi-Domain Sentiment Analysis with Domain Adaptation Using BERT,” Information Sciences, vol. 546,
40. pp. 451–463, 2021.
41. S. Chatterjee and M. Narayan, “Sentiment Classification on IMDb Movie Reviews Using Transformer Models,” Procedia Computer Science, vol. 218, pp. 702–709, 2023.
42. K. Choudhary and A. Jain, “Comparative Study of NLP Models for Text Sentiment Classification,” Springer Advances in Intelligent Systems and Computing, vol. 1389, pp. 125–136, 2022.
43. D. Jurafsky and J. H. Martin, Speech and Language Processing (3rd Edition Draft), Prentice Hall, 2020.