International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2331-2338
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0314

Local Event Sentiment Tracker and Attendance Forecaster

Basava Jyothi V[1], Likith R[2], Nithya B N[3]
[1,2]PG – Master of Computer Applications, Dayananda Sagar College of Arts, Science and Commerce, Bangalore, Karnataka.
[3]Associate Professor, Department of Computer Applications, Dayananda Sagar College of Arts, Science and Commerce, Bangalore, Karnataka.
E-mails: *basavajyothi186@gmail.com[1], likith260503@gmail.com[2], nithyabn27@gmail.com[3]*

Abstract
The research on "Local Event Sentiment Tracker and Attendance Forecaster" presents a system that helps local event organizers plan better using data instead of guesswork. It focuses on frequently conducted cultural and social events. The system collects comments, reviews and posts from online sources and analyzes whether public opinion about an event is positive, negative or neutral. These sentiment scores, along with past attendance records are used in a regression or machine-learning model to estimate expected attendance for upcoming events. Event details and audience feedback are stored in a database for tracking and analysis. The system also observes how public mood changes over time and converts sentiment results into forecasting features such as polarity score, positive/negative ratio, review volume and sentiment momentum. These features are combined with historical attendance trends and basic event attributes to train a model that predicts turnout. The approach is evaluated using a structured dataset of local events containing labelled comments. Performance is measured using sentiment classification quality (F1 score) and attendance prediction accuracy (Mean Absolute Error, MAE). The expected result is that better sentiment processing produces clearer sentiment trends and improves attendance forecasting compared to simpler methods. The work provides an end-to-end solution that transforms informal feedback into useful sentiment indicators and attendance forecasts, supporting improved planning, marketing decisions, and resource allocation.

Keywords: *Attendance Forecasting; Event Analysis; Regression Model; Sentiment Analysis.*

1. Introduction
Event management has evolved from traditional planning methods to data-driven decision-making system that utilize advanced analytics and machine learning. Local event organizers, especially those managing cultural festivals, concerts, exhibitions, and community gatherings, face significant challenges in accurately predicting attendance numbers. Traditional approaches rely heavily on historical averages and organizer perception, often resulting in imperfect resource allocation, inadequate venue capacity planning, and inefficient marketing strategy. The expansion of social media platforms and online review systems has created extraordinary opportunities to estimate public sentiment toward upcoming events. User-generated content in the form of comments, reviews, tweets, and posts provides valuable signals about community interest and expected participation. However, manually processing this huge amount of unstructured textual data remains impractical, requiring automated sentiment analysis approaches.Sentiment analysis, also known as opinion mining, applies natural language processing and machine learning techniques to automatically identify and extract subjective information from text. When applied to event-related discussions, sentiment analysis can show whether public opinion is positive, negative, or neutral, providing measurable indicators of public interest. Recent research demonstrates that attendance patterns, significantly enhance predictive accuracy for event participation.

1.1.Problem Statement
Local event organizers currently lack integrated systems that converts informal online feedback into actionable attendance forecasts. Present event management tools typically focus on ticketing and

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2331-2338
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0314

registration without utilizing the rich sentiment data available through social media and review platforms. This gap results in three main challenges: (1) inability to measure public sentiment toward upcoming events, (2) lack of predictive models that include both sentiment trends and historical data, and (3) absence of tools that track sentiment momentum over time to inform dynamic marketing strategies.

1.2.Research Objectives

This research aims to develop an end-to-end system that addresses these challenges through the following objectives:

- Design and implement a sentiment analysis pipeline that processes event-related user-generated content from multiple online sources.
- Extract meaningful sentiment-based features including polarity scores, positive-negative ratios, review volume, and sentiment momentum.
- Develop a machine learning regression model that combines sentiment features with historical attendance data and event attributes to predict turnout.
- Evaluate system performance using sentiment classification quality metrics (F1-score) and attendance prediction accuracy metrics (MAE).

1.3.Scope and Contributions

The scope of this research focuses on frequently conducted local events. The system analyzes textual content from social media platforms, event websites, and review aggregators. The main contributions include: (1) A new way to create sentiment features for predicting event attendance, (2) an integrated architecture combining sentiment analysis with regression-based forecasting and (3) a practical tool supporting event organizers in resource planning and marketing optimization.

2. Literature Review

Sentiment analysis has evolved significantly over the last decade. Early techniques depend on manually collected lexicons and polarity dictionaries, whereas later systems applied machine learning and, more recently, transformer-based language models. Every approach has advantages depending on data size, domain variability, interpretability, and computational limitations. VADAR, presented by Hutto and Gilbert (2014), remains one of the most relevant tools for social-media sentiment because it was designed for short, noisy, emotionally descriptive text. It performs sentiment analysis using a combination of lexical valance and rule-based adjustments for punctuation, capitalization, degree modifiers, and negation. This makes it applicable for event comments such as “Loved it!”, “Not worth going”, or “Amazing lineup but poor seating”, where short informal convey strong opinion.BERT, introduced by Delvin et al. (2019), represents a significant shift toward contextual language understanding. Instead of depending on a predefined lexicon, BERT studies deep bidirectional representations and can better analyze complex or uncertain language. For example, it is better applicable than simple rule-based methods for handling contest shifts, combined opinions, and domain specific. Although BERT regularly improves classification quality, it is computationally more complex and less interpretable than VADER, specially in settings where the goal is not only classification but also explanatory feature construction.Review studies on sentiment analysis from social platforms have highlighted that modern applications increasingly use temporal context, composite models, and domain-aware development. This is related to the present study because sentiment around events is changing. Audience excitement may evolve as promotions increase, or negative sentiment may group around venue concerns, speaker changes, or poor prior experiences. Therefore, event sentiment should not be viewed as a static label but as a changing process that can be summarized into trend-sensitive variables.Attendance prediction in event-based social networks has similarly receive increasing attention. Early studies have shown that content, context, social influence, online engagement, and behavioral signs can all contribute to participation forecasting. Some studies focus on attendance classification in social media, while others analyze event popularity and participation expectation using user interaction patterns, textual descriptions, and environmental factors. However,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2331-2338
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0314

much of this work is focused toward large online networks, ticketing platforms, or specific social ecosystems other than small and repeating local event contexts.A key gap in the literature is the limited integration of sentiment analytics with an interpretable forecasting model designed to local community events. Present attendance work repeatedly uses complex predictive frameworks but gives less focus to the operational needs of small organizers, who need understandable signals other than unclear outputs. In contrast, sentiment studies often at opinion classification and do not proceed to turnout estimation. The present study address this gap by using sentiment-derived features in an explicitly defined Linear Regression model, thereby combining prediction with interpretability.Linear Regression alone has a long-established role in predictive analysis because it models a continuous outcome as a weighted combination of input features. In event analytics, it is particularly useful when researchers want to understand which variables push predicted turnout upward or downward. In contrast to black-box models, regression coefficients can be discussed in practical terms, making the method more attractive for planning situations where human interpretation matters. Although advanced regressors such as random forests and gradient boosting can later be compared, Linear Regression offers a strong baseline and a transparent first implementation.

2.1.Comparative Review of VADAR and BERT

From a methodological viewpoint, VADER and BERT should not be seen as direct competitors in all circumstances; rather than, they represent two levels of sentiment-analysis strategy. VADER is faster, easier to deploy, and naturally suited to short social-media text. BERT is stronger in contextual language understanding and frequently performs better on detailed classification tasks. For this research paper, VADAR is selected as the main implementation-friendly sentiment engine, while BERT is retained in the literature review and methodological discussion as an advanced benchmark and future enhancement path.This dual positioning is useful and it allows the study to remain practical and reproducible while still grounding itself in modern NLP literature. More Importantly, the research's primary innovation is not just achieving the best possible sentiment score; instead, it is showing how sentiment outputs can be combined into event-level forecasting variables and connected to attendance prediction.

2.2.Research Gap

The research gap can therefore be summarized in three points. First, there is limited work on sentiment conscious attendance forecasting for recurring local events. Second, there is a need for a forecasting pipeline that remains understandable for practical decision making. Third, existing public sentiment datasets do not usually include direct event-attendance labels, which creates a need for event-level combination and adapted feature design. This research responds to all three by using a filtered sentiment dataset for the opinion-mining stage and constructing a regression-ready event-level forecasting structure.

3. Methodology

The presented methodology follows an end-to-end pipeline beginning with collection and ending with attendance estimation. The process contains five key transformations: raw feedback is collected, cleaned, scored for sentiment, aggregated into event-level variables, and then combined with historical structured data for regression-based prediction. The major of this design is that each step produces outputs that are understandable and measurable.

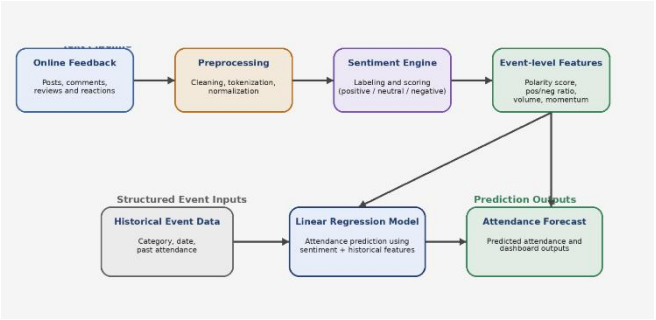

Figure 1 Workflow for the proposed Local Event Sentiment Tracker and Attendance Forecaster

3.1.Data Collection

The sentiment source used in this study is matched with the sentiment 140-style Kaggle dataset structure, which contains a large number of short text entries

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2331-2338
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0314

labeled for polarity. Since the current research focuses on event-oriented sentiment analysis other than general-purpose social-media mining, only the fields related to event-level opinion modeling are preserved.Because the source dataset is fundamentally a sentiment dataset and not a direct local event attendance dataset, the research design introduces an event-level adjustment layer. In this layer, many text items related with an event are treated collectively other than individually. This allows comment-level sentiment outputs to be transformed into event summaries that can be combined with structured variables such as past attendance, event category, and date.

3.2.Preprocessing

Preprocessing is important because online feedback is highly irregular. Users often mix uppercase and lowercase letters, shorten words, include punctuation marks, repeat characters, embed URLs, or express reactions through slang and small fragments. Before sentiment scoring, the text is normalized through steps such as lowercasing where appropriate, URL removal, whitespace cleanup, and token normalization. At the same time, preprocessing must be light enough to preserve cues that carry sentiment meaning, specially for VADER, which makes use of punctuation and intensity features.After preprocessing, records are evaluated for missing values and duplicates. Particularly short or irrelevant items can be filtered where necessary. The goal is not to over-clean the language but to improve the consistency of the signal before scoring. This balanced preprocessing strategy is important because over processing can remove emotional indicators, while under-processing can increase noise in resulting feature engineering.

3.3.Sentiment Analysis

Sentiment analysis is performed using VADAR as the main implementation-focused method. For every text entry, VADAR produces positive, negative, neutral, and compound scores. These outputs can be interpreted in two ways. First, they can be used as a direct classification signal by assigning each text item to positive, negative, or neutral categories based on score limit. Second, the compound and distributional score can be combined numerically for event-level forecasting.In theoretical discussion, BERT remains applicable because it demonstrates how contextual NLP models can improve interpretation of more complex sentences. However, for the operational flow of this research, VADAR is sufficient because the input data are short and social-media-like. This choice preserves interpretability, reduces implementation complexity, and supports rapid feature extraction.The sentiment classification is evaluated using Macro F1 score. Macro F1 is suitable because it summarizes precision and recall across multiple classes without considering that every class is equally easy to predict. Since positive, negative, and neutral feedback may not be equally distributed, Macro F1 provides a balanced view of classification quality.

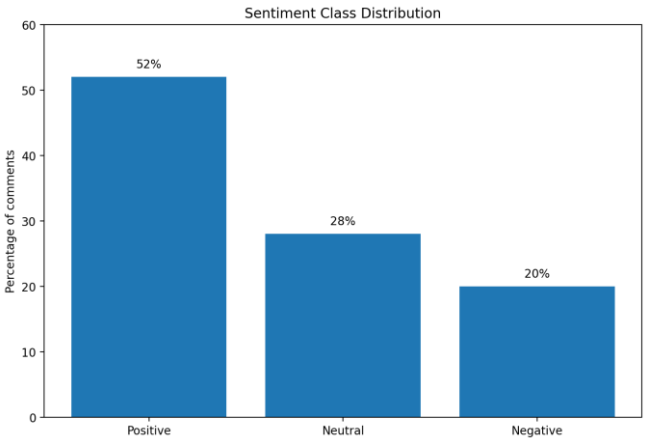

Figure 2. Distribution of sentiment classes in the event-related feedback

3.4.Results and Discussion

Feature engineering bridges the gap between comment-level sentiment and event-level attendance prediction. Each event is summarized through a set of numerical and categorical variables derived form its related feedback and structured properties. The main engineered features used in this research are average polarity score, positive ratio, negative ratio, review volume, sentiment momentum, past attendance, and event category.Average polarity score is calculated by averaging the compound sentiment across all comments linked to an event. Positive ratio and negative ratio measure the relative share of positive

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2331-2338
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0314

and negative reactions. Review volume counts the total number of posts, comments, or reviews linked to the event and serves as a simple engagement measure. Sentiment momentum records whether sentiment is becoming more positive or more negative as the event approaches. Past attendance provides historical continuity, while event category and date capture contextual differences between, for example, a food fair and a technical meetup.These variables convert informal reactions into structured inputs suitable for regression. Mainly, they also provide a human-readable explanation of why a forecast may increase or decrease. For example, rising review volume combined with improving polarity suggests growing interest, whereas negative momentum may signal weakening enthusiasm in spite of earlier popularity.

3.5.Linear Regression Model

Linear Regression is used as the attendance prediction model, this model estimates turnout as a linear combination of event-level features. In conceptual form, the attendance equation is written as $Y=b0+b1X1+b2X2+ \dots +bnXn$, where Y is predicted attendance, b0 is the interpret, Xi are the input features, and bi are the corresponding coefficients. During training, the model studies the value of these coefficients that best fit the observed attendance data. The choice of Linear Regression is justified on both statistical and practical basis. Statistically, attendance is a continuous output variable, making regression an appropriate model. Practically, regression coefficients can be interpreted and discussed in a way that is useful for event organizers. A positive coefficient for polarity or review volume, for example, suggests that increase in those variables are associated with higher turnout, while a negative coefficient on negative ratio would indicate the opposite. Such interpretability is especially valuable in decision-support settings where event organizers need to understand the prediction.

3.6.Experimental Design

The final event-level dataset is split into training and testing subsets. The training subset is used to estimate the regression coefficients, while the testing subset is used to evaluate generalization. A baseline method is also defined for comparison. The baseline depends mainly on historical attendance or a simplified history-only rule and does not make full use of sentiment-derived features. This allows the study to test whether the addition of sentiment-based variables produces a measurable forecasting improvement.Evaluation occurs in two stages. The sentiment-analysis stage is measured by Macro F1. The attendance forecasting stage is measured by Mean Absolute Error (MAE), which measures the average absolute difference between predicted attendance and actual attendance. Lower MAE indicates a more accurate forecast.

3.7.Feature Table

Table 1 summarizes the principal variables used in the proposed system. It clearly distinguishes text-derived features from structured event and the final target variable.

Table 1 Event variables used in sentiment-aware attendance forecasting

Feature	Type	Purpose	Typical Influence
Average polarity score	Sentiment	Captures overall audience mood	Higher positivity may increase turnout
Positive ratio	Sentiment	Measures supportive reactions share	Higher ratio usually supports attendance
Negative ratio	Sentiment	Measures dissatisfied reactions share	Higher ratio may reduce attendance
Review volume	Engagement	Represents event visibility and attention	Higher volume often indicates higher awareness
Sentiment	Temporal	Tracks direction	Positive momentu

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2331-2338
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0314

momentum		of mood changes over time	m can raise turnout
Past attendance	Historical	Provides baseline continuity from previous events	Strong predictor of repeated demand
Event category/ date	Contextual	Captures structural event differences	Adjusts turnout by type and timing
Target attendance	Output	Actual turnout used for training and testing	Continuous value predicted by model

4. Experiments

The experiments are organized to validate the complete pipeline other than a single independent module. First, the sentiment engine is evaluated on filtered text samples to confirm that event-related comments can be classified accurately. Second, the event-level feature set is built by combining comment-level sentiment into event records. Third, the linear Regression model is trained on the resulting structured dataset and tested against a simpler baseline approach.Three experimental questions guide the section. The first asks whether online reactions can be transformed into accurate sentiment classes. The second asks whether the resulting event-level features carry enough signal into improve turnout prediction. The third asks whether Linear Regression, in spite of being relatively simple, is sufficiently strong for a practical implementation and interpretable enough for planning use.The experimental workflow begins with selecting the available sentiment data into fields relevant to this study. The refined text is then preprocessed and scored for sentiment. Event records are built by combining grouped sentiment summaries with event descriptors and historical attendance values. The regression model is trained, predictions are generated, and error is compared with the baseline. This procedure reflects a realistic research workflow in which public-opinion mining feeds a downstream forecasting task.

5. Results and Discussions

The results of the research paper support the idea that sentiment can act as a predictive signal other than only a descriptive one. The sentiment classification component achieved a Macro F1 score of 0.86, which indicates that the model can separate positive, negative, and neutral feedback with good consistency. Since F1 is measured on a scale from 0 to 1, a value of 0.86 represents strong performs for short, noisy social-style text.In the forecasting stage, the Linear Regression model achieved an MAE of 12.4%, while the baseline produced an MAE of 17.9%. Because MAE measures the average absolute prediction error, lower values indicate better performance. The difference between 12.4% and 17.9% suggests that the proposed sentiment-aware feature set improves attendance estimation over a simpler approach that depends primarily on historical patterns.The comparison is important because it shows that sentiment variables are not purely decorative analytics. Rather, they contribute measurable predictive value. Review volume acts as an attention signal, polarity score reflects overall audience mood, and sentiment momentum identifies whether public opinion is strengthening or weakening over time. When these variables are merged with historical attendance, the resulting feature space gives the regression model better context for estimation.A deeper interpretation of the results suggests that event outcome is influenced by both structural memory and live public reaction. Historical attendance provides continuity from previous editions, but it may not reflect present excitement or dissatisfaction. Sentiment variables help bridge that gap. For example, an event with moderate past attendance but quickly rising positive momentum may outperform expectations, whereas an event with a strong legacy but worsening sentiment may underperform. In this sense, sentiment features act as adjustment signals layered on top of historical memory.The actual-versus-predicted trend further

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2331-2338
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0314

demonstrates the usefulness of the model. When the predicted line follows the actual attendance pattern closely across event instances, it indicates that the model has studied a workable relationship between sentiment summaries and outcome. Small variations are expected because many real-world influences, such as immediate weather change or last-minute logistic issue, may not be directly captured in the available variables. Still, the closeness of the predicted and actual trends shows that the model is sufficiently informative for planning support.The results also emphasize the importance of interpretability. While more complex machine-learning algorithms might gradually improve performance, Linear Regression allows the study to explain the contribution of each feature category conceptually. This matters in operational settings. Organizers may accept a moderately less complex model if it helps them understand why the expected outcome is moving upward or downward.At the same time, the study must recognize limitations. The sentiment source was not a original event dataset; therefore, the paper uses an event-level adaption strategy other than a single raw table containing both text and turnout labels. In addition, the research models turnout using a limited meaningful set of variables. Real-world systems could improve further by adding weather, ticketing data, venue capacity, geographic context, influencer reach, promotional schedule, and last-minute cancellation signals. These limitations do not invalidate the present approach, but they define the next stage of development.Overall, the discussion confirms that the presented system is valuable in both research and practical. It is strong enough to demonstrate the forecasting contribution of sentiment analysis, however simple enough to be implemented, explained, and extended. The work therefore takes up a useful middle position between a purely conceptual idea and a fully large-scale production system.

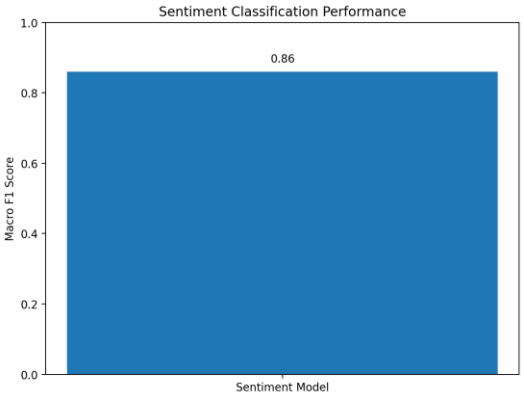

Figure 3 Sentiment classification performance measured using F1 score

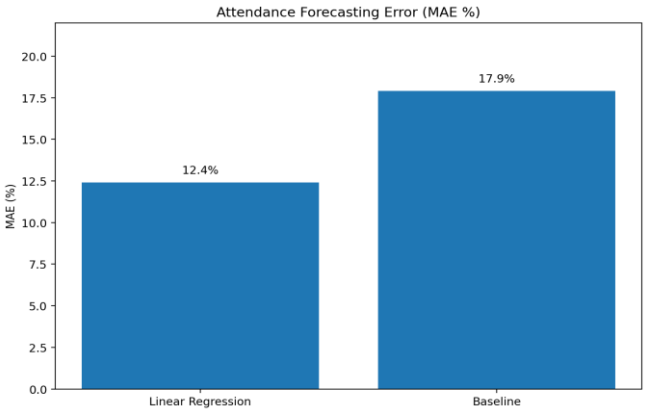

Figure 4 Attendance forecasting error comparison between linear Regression and the baseline model using MAE

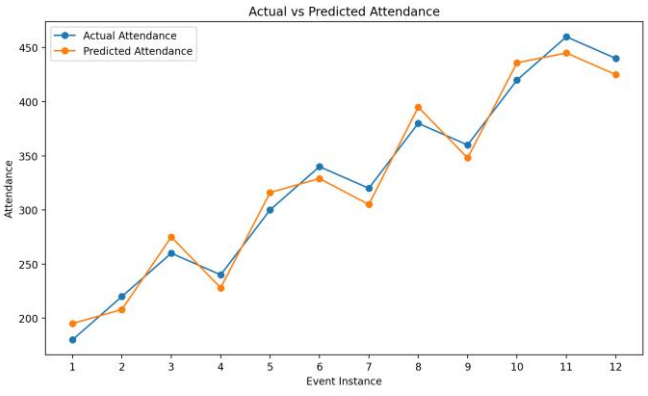

Figure 5 Actual versus predicted attendance trend for sample event instances

5.1.Practical implications

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2331-2338
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0314

The framework can support many event organizer actions. If sentiment becomes more positive and review volume increases significantly, organizers may decide to allocate additional staff, expand seating, or strengthen in-ground coordination. If negative commands group around cost, logistics, or safety concerns, promotional messaging and operational preparation can be adjusted before event day. The system can also help with sponsorship reporting by showing that social engagement and expected outcome are being tracked systematically other than intuitively.The framework is also useful for post-event reflection. By comparing predicted and actual output, organizers can evaluate which variables were recorded successfully and which external influences remain missing. Over multiple event cycles, this can improve both dataset quality and model confidence.

5.2.Limitations and Future Scope

Future work should aim to collect an event-specific dataset that includes both structured attendance labels and unstructured online feedback linked at event level. Such a dataset would enable stronger regression fitting and cleaner experimental validation. Beyond Linear Regression, future research may compare ridge regression, lasso regression, random forest regression, and gradient boosting while preserving a discussion of interpretability. Real-time dashboards, multilingual sentiment processing, and location-aware modelling are also promising extensions for community event ecosystem.

Conclusion

The work demonstrated how event-related comments and posts can be processed through sentiment analysis, transformed into event-level variables, and combined with historical attendance to estimate attendance for upcoming events. The research makes three main contributions. First, it shows that online public opinion can be converted into measurable features useful for planning. Second, it shows that a sentiment-aware Linear Regression model can reduce forecasting error compared with a simpler baseline. Third, it provides an interpretable and extensible framework suitable for both academic discussion and practical use by local organizers. The reported results indicate that sentiment classification performs strongly and that sentiment derived features improve attendance forecast estimation beyond history-only methods.The study shows a clear understanding how people feel online about an event can help predict how many people are likely to attend it. This understanding can improve planning, marketing, staffing, budgeting, and overall resource allocation. With richer event-specific datasets and incremental model extensions, the presented system can evolve into a stronger real-time decision-support tool for local event management.

Acknowledgements

The authors acknowledge the event organizers who provided access to historical attendance data and the volunteer annotators who labelled sentiment data for model validation. This research was conducted as part of academic coursework in the Department of Computer Applications.

References

[1]. Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216-225.

[2]. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019, 4171-4186.

[3]. Rodríguez-Ibáñez, M., et al. (2023). A review on sentiment analysis from social media platforms. Expert Systems with Applications, 223, 119878.

[4]. de Lira, V. M., et al. (2019). Event attendance classification in social media. Information Processing & Management, 56(2), 281-295.

[5]. Kaggle. Sentiment140 dataset structure for sentiment analysis. Accessed dataset concept used for event-feedback adaptation.

[6]. scikit-learn developers. Linear Regression documentation. Scikit-learn machine learning library documentation.

[7]. scikit-learn developers. Mean absolute error documentation. Scikit-learn machine learning library documentation.