Indonesian Journal of Electrical Engineering and Computer Science
Vol. 39, No. 1, July 2025, pp. 353~363
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v39.i1.pp353-363

Enhancing TV program success prediction using machine learning by integrating people meter audience metrics with digital engagement metrics

Khalid El Fayq, Said Tkatek, Lahcen Idouglid
Laboratory of Computer Sciences Research, Faculty of Science, Ibn Tofail University, Kenitra, Morocco

Article Info

Article history:

Received Aug 27, 2024
Revised Mar 11, 2025
Accepted Mar 25, 2025

Keywords:

Audience reaches
Data regression models
Digital engagement
Media analytics
Television audience prediction
Watermarking audience

ABSTRACT

With the emergence of numerous media services on the internet, television (TV) remains a highly demanded medium in terms of reliability and innovation, despite intense competition that compels us to devise strategies for maintaining audience engagement. A key indicator of a TV channel's success is its reach, representing the percentage of the target audience that views the broadcasts. To aid TV channel managers, the industry is exploring new methods to predict TV reach with greater accuracy. This paper investigates the potential of advanced machine learning models in predicting TV program success by integrating people meter audience metrics with digital engagement metrics. Our approach combines convolutional neural networks (CNNs) for processing digital engagement data, long short-term memory (LSTM) networks for capturing temporal dependencies, and gaussian processes (GPs) for modeling uncertainties. Our results demonstrate that the best-performing hybrid model achieves a prediction accuracy of 95%. This study contributes to the field by addressing manual scheduling errors, financial losses, and decreased viewership, providing a more comprehensive understanding of audience behavior and enhancing predictive accuracy through the integration of diverse data sources and advanced machine learning techniques.

Corresponding Author:

Khalid El Fayq
Laboratory of Computer Sciences Research, Faculty of Science, Ibn Tofail University
Kenitra, Morocco
Email: khalidelfayq@gmail.com

1. INTRODUCTION

The marketing industry is one of the largest globally, with television (TV) companies investing millions of dollars in advertising. Audience ratings play a pivotal role in guiding advertisers to tailor campaigns and helping content creators and TV networks refine programming strategies to maximize audience engagement and optimize content promotion. In Morocco, audience measurement has achieved significant advancements, spearheaded by organizations such as the interprofessional media audience center (CIAUMED) and Marocmetrie. Using techniques like watermarking and fingerprinting, Marocmetrie monitors TV consumption for over 1,000 households, providing valuable insights into audience behavior. Despite these advancements, challenges persist-particularly for Laayoune TV, where manual scheduling of daily TV guides often results in misjudged demographic data, inaccurate TV ratings, and suboptimal advertisement pricing. These inefficiencies can lead to financial losses, decreased advertising effectiveness, and diminished viewership, emphasizing the necessity of an automated, data-driven approach for predicting TV program success.

To address these challenges, this study introduces a novel predictive framework that integrates people meter audience metrics with digital engagement metrics to improve the accuracy of broadcast scheduling and program success forecasting. Unlike traditional methods that rely solely on audience panel data, our approach incorporates advanced machine learning models to analyze and synchronize multiple data sources, capturing both real-time TV audience behavior and digital interactions, specifically from YouTube. Specifically, this study leverages a hybrid machine learning architecture that combines convolutional neural networks (CNNs) to extract deep representations from digital engagement data, such as user interactions, comments, and content shares on YouTube, long short-term memory (LSTM) networks, a type of recurrent neural network (RNN), to effectively model temporal dependencies in audience behavior over time, capturing fluctuations in TV viewership patterns, and gaussian processes (GPs) to account for uncertainties and variations in audience metrics, allowing for a more robust and interpretable prediction framework. This multi-layered approach enhances predictive accuracy by integrating diverse structured data (TV ratings, scheduling metadata) and unstructured data (digital engagement metrics), enabling a more adaptive modeling process compared to static panel-based audience measurement methods. By reducing manual scheduling errors and optimizing advertising placements, this integration facilitates improved decision-making for content scheduling while ensuring greater resilience to market fluctuations, as the inclusion of real-time digital engagement data allows for rapid adaptation to shifts in audience preferences.

Previous studies, such as those conducted by Nixon [1], Cammarano *et al.* [2], and Lucas and Lazarus [3], have explored audience prediction using machine learning, they primarily focused on single data sources or limited metrics. Our previous work [4] utilized only people meter audience metrics, demonstrating the potential of machine learning for program prediction but lacking the integration of digital engagement data, a critical component in modern audience measurement. By bridging this gap, the current study advances predictive analytics for TV programming.

Recent works, such as Zhou *et al*. [5] and Jeyavadhanam *et al.* [6], have demonstrated the utility of machine learning in predicting online TV video success. Similarly, Verma [7], Oyewola and Dada [8] and Gupta *et al.* [9] investigated models for movie success prediction. Studies by Abarja [10], Sharma *et al.* [11], Cizmeci and Oguducu [12], and Crisci *et al.* [13] highlighted the role of social media metrics in understanding audience behavior, while Kupavskii *et al.* [14] underscored the importance of integrating social and traditional metrics. However, these approaches often lack the integration of temporal and uncertainty modeling, as required for dynamic audience prediction.

Studies by Song *et al.* [15] and Akgül and Küçükyilmaz [16] utilized aggregated people meter data to forecast TV ratings. Advanced models, including neural networks [17], [18], gradient-boosting machines [19], ridge regression by Ma *et al.* [20], Seric *et al.* [21], and Choi *et al.* [18], decision trees [22], and genetic algorithms by Gegres *et al.* [23] have been applied to optimize accuracy. These approaches, however, often lack the integration of temporal and uncertainty modeling, which are critical for dynamic audience behavior.

This study builds upon prior works by integrating people meter audience metrics with digital engagement metrics, addressing limitations in siloed approaches. We employ random forest (RF) [24], K-nearest neighbors (KNN) [25], and advanced models such as CNNs, LSTM networks [26], effectively capture temporal dependencies in viewing patterns, while GPs model uncertainties, enabling robust predictions. By bridging the gap between traditional and digital metrics, this study offers a novel approach to enhancing TV program success prediction.

The remainder of this paper is structured as follows: section 2 describes the proposed methodology, covering data acquisition, preprocessing, algorithm selection, performance evaluation, deployment, and recommendations. Section 3 presents the experimental results and discussion, including model performance, comparative analysis, case studies, and an in-depth discussion of the findings. Finally, section 4 concludes with key insights and potential directions for future research.

2. METHOD

In this section, we describe the methodology adopted in this study, detailing the overall system architecture, data acquisition process, preprocessing steps, model selection, evaluation, and deployment strategies. The proposed framework integrates people meter audience metrics with digital engagement metrics to enhance TV program success prediction through machine learning techniques. Figure 1 illustrates the dynamic system architecture, which outlines the end-to-end process, from data acquisition to model deployment. The system begins with data collection from two primary sources-people meter audience metrics from CIAUMED and YouTube engagement data-followed by data preprocessing, feature extraction, and model training. The predictions generated by the model are then used for content scheduling recommendations.

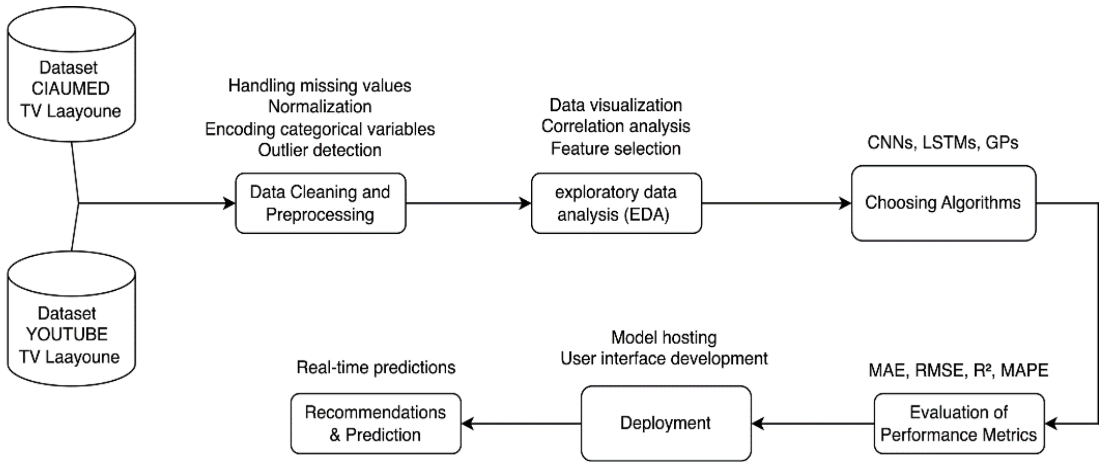

Figure 1. Dynamic proposed system architecture

2.1. Data acquisition

To develop a robust prediction model, data were collected from two primary sources. The first source is the people meter audience metrics, collected by CIAUMED, this data provides detailed statistics on traditional TV viewership across Morocco. Metrics such as the number of viewers, program duration, and audience demographics were captured using people meters installed in a representative sample of over 1,000 households.

The second source, digital engagement metrics, comprises data from 3,500 videos collected from the YouTube channel of TV Laayoune. This dataset includes views, likes, comments, and engagement rates for online content. The integration of these two data sources is essential to bridge the gap between traditional TV audience measurement and online digital consumption trends. Figure 2 presents the process of combining traditional TV success metrics with online video engagement, demonstrating how these datasets are aligned to enhance predictive modeling.

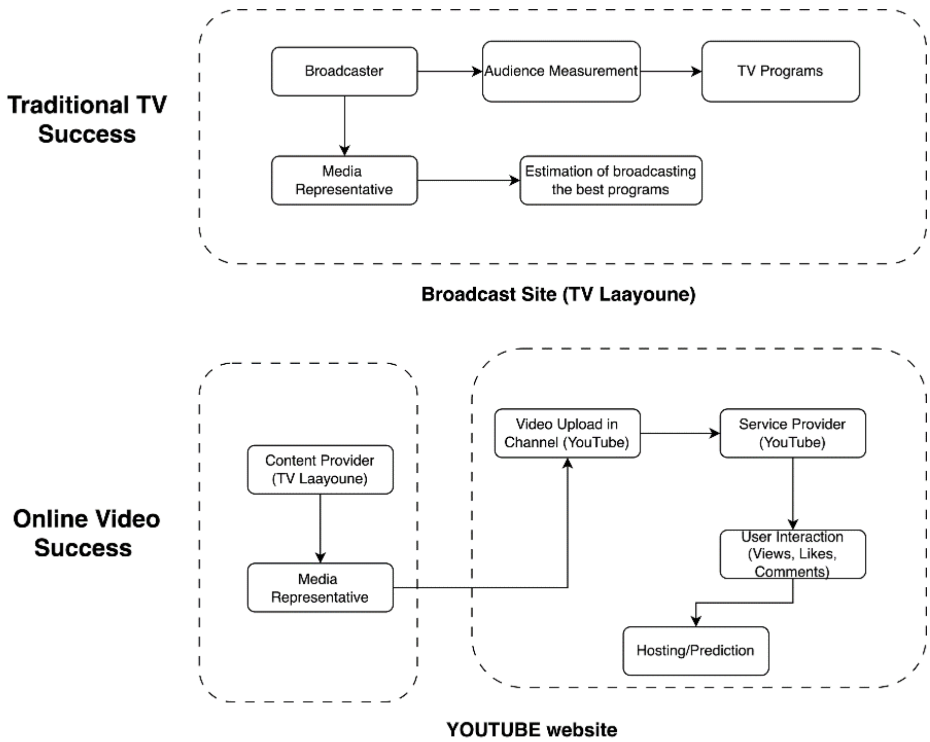

Figure 2. Process of traditional TV success and online video success

2.2. Data cleaning and preprocessing

Data preprocessing is a critical step to ensure the quality and reliability of the input data used for model development. It involves several stages to prepare and standardize the data for optimal performance in machine learning algorithms. The first step involves cleaning the raw data from both people meter audience metrics and digital engagement metrics to remove inconsistencies and errors. Missing values are handled through imputation techniques, where missing data points are estimated using statistical methods or values derived from similar records in the dataset. This ensures that incomplete data does not negatively impact the model's performance.

Numerical features are normalized to bring their values within a comparable range, ensuring uniformity across different scales. This normalization is achieved using the (1).

$$x' = \frac{x - \min(x)}{max(x) - \min(x)} \tag{1}$$

Where x is the original value, x' is the normalized value, $\min(x)$ is the minimum value, and $\max(x)$ is the maximum value in the dataset.

Categorical variables were encoded into numeric representations using one-hot encoding, which converts categories into binary vector representations, ensuring compatibility with machine learning algorithms. Outliers were identified and addressed using statistical methods such as the Z-score, which detects values significantly outside the normal range. These preprocessing steps ensured the data's consistency, accuracy, and compatibility with machine learning models.

Finally, exploratory data analysis (EDA) is conducted to understand underlying patterns and relationships in the data. This includes visualizing distributions, correlations, and identifying significant features that will inform the choice of algorithms for the model. By following these preprocessing steps, the dataset is prepared to ensure consistency, accuracy, and compatibility with machine learning algorithms, ultimately improving the reliability and robustness of the predictive model.

2.3. Choosing algorithms

The primary objective of this study is to develop a predictive model that identifies TV programs likely to attract the largest audience. To achieve this, we selected a combination of machine learning algorithms that capture diverse aspects of the dataset, such as spatial patterns, temporal dependencies, and uncertainties, ensuring comprehensive and accurate predictions. We evaluated various machine learning algorithms to identify the most suitable ones for predicting TV program success. The algorithms chosen includes:

- RF: was selected for its ability to handle high-dimensional data effectively while minimizing the risk of overfitting. Its ensemble learning approach combines multiple decision trees to provide stable and interpretable predictions. RF is particularly suitable for capturing relationships in people meter audience metrics, where categorical and numerical data are abundant.
- KNN: was chosen for its simplicity and ability to classify data points based on local patterns. It is effective in capturing small-scale trends within the data, such as viewer preferences for specific content categories.
- Support vector machine (SVM): was selected for its robustness in high-dimensional spaces. Its ability to find the optimal hyperplane for separating classes makes it a strong candidate for predicting whether a TV show will succeed based on mixed numerical and categorical data.

While traditional algorithms like RF, KNN, and SVM provide a good foundation, they lack the ability to effectively capture temporal trends and uncertainties, which are critical for predicting dynamic viewership behavior. To address the limitations of traditional algorithms, this study integrates advanced deep learning and probabilistic methods into a hybrid model:

- Convolutional neural networks (CNNs): are integrated into the hybrid model to process the digital engagement metrics dataset from YouTube. These metrics often have a spatial structure (e.g., sequential data or video-specific patterns), and CNNs excel at recognizing such spatial relationships. By extracting high-level features from engagement data (e.g., likes, comments, shares), CNNs help identify patterns that correlate with TV show popularity.
- Long short-term memory (LSTM): are essential for capturing temporal dependencies in the data. Viewer behavior often follows temporal patterns-specific times of day, days of the week, or seasonal trends influence TV viewership. LSTMs are particularly suited for modeling such sequential dependencies, ensuring the model can predict viewership trends over time.
- GPs: were incorporated into the model to handle uncertainties in predictions. Unlike other algorithms, GPs not only provide a point estimate but also quantify the confidence of the prediction. This is crucial

when recommending a TV show for optimal scheduling, as it allows decision-makers to understand the potential risks associated with each prediction.

The integration of CNNs, LSTMs, and GPs into a single hybrid model leverages the strengths of each algorithm, ensuring that this hybrid approach delivers more accurate and reliable prediction of TV show success than standalone models. Figure 3 illustrates the integration of CNNs, LSTMs, and GPs.

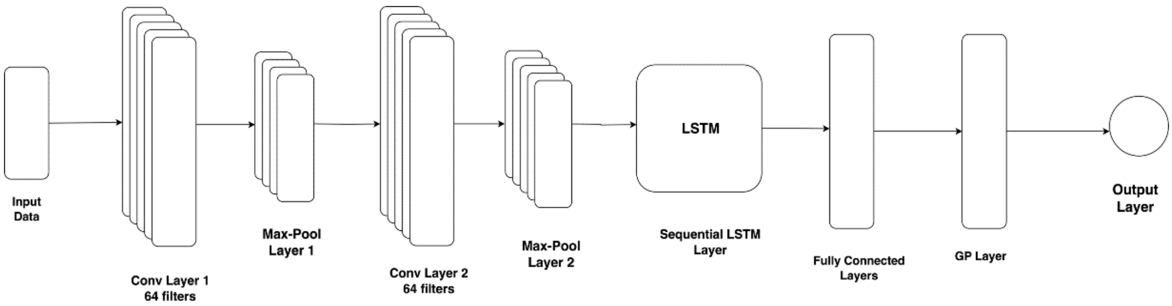

Figure 3. Machine learning model architecture for predicting TV program success

Table 1 outlines the hyperparameters governing our combined CNN, LSTM, and GP architecture. These hyperparameters dictate the architecture and training of the CNN responsible for feature extraction, the LSTM network for capturing temporal dependencies, and the GPs for modeling uncertainties. This approach ensures the effective extraction of informative features, capturing temporal patterns, and modeling uncertainties from the input data, leading to superior performance in predicting TV program success.

Table 1. Hyperparameters for CNN, LSTM, and GP architecture

Learning rate	Batch size	Number of epochs	Dropout rate	Kernel size	Filters	Activation function	Weight initialization
0.001	32	30	0.5	3×3	64,128	Rectified linear unit (ReLU)	He initialization

2.4. Evaluation of performance metrics

The performance of the model was evaluated using multiple metrics to ensure a comprehensive assessment of its predictive capabilities. These metrics include mean absolute error (MAE) (2), root mean squared error (RMSE) (3), R-squared (R^2) (4), and mean absolute percentage error (MAPE) (5). Each metric provides unique insights into the model’s performance, allowing us to measure both accuracy and robustness. The performance formulae are given below.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (2)$$

Where n is the number of observations, y_i is the actual value, and $\hat{y}_i$ is the predicted value. MAE measures the average magnitude of errors in a set of predictions, without considering their direction.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

RMSE provides a measure of the differences between values predicted by a model and the values observed, with a higher penalty on larger errors.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (4)$$

Where $\bar{y}$ is the mean of the actual values. R^2 indicates the proportion of the variance in the dependent variable that is predictable from the independent variables, providing insight into the goodness-of-fit of the model.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (5)$$

MAPE measures the accuracy of a forecasting method as a percentage, offering an intuitive understanding of the prediction errors relative to the actual values. These metrics collectively provide a comprehensive assessment of the model's predictive accuracy, robustness, and generalizability.

2.5. Deployment

The deployment phase integrates the predictive model into TV Laayoune's workflow to provide real-time insights for programming decisions. Hosted on a server using TensorFlow Serving, the model is accessible via REST API endpoints, ensuring seamless integration with existing infrastructure and high-performance predictions. A real-time data pipeline continuously feeds new people meter audience metrics and digital engagement metrics into the model, preprocessing the data to ensure consistency and reliability. The model is embedded into the scheduling and broadcasting systems, automating recommendations for optimal broadcast times and reducing reliance on manual scheduling.

Figure 4 illustrates the interface web for predicting TV program success. The system allows users to select a specific TV program, view its audience performance on TV, and compare it with its digital footprint on YouTube. By integrating these insights, decision-makers can optimize scheduling strategies and advertising placements. To maintain accuracy, the model undergoes periodic retraining by incorporating updated audience data. This allows the system to adapt to changing viewer behaviors, seasonal trends, and evolving content preferences, ensuring continuous improvement in prediction accuracy.

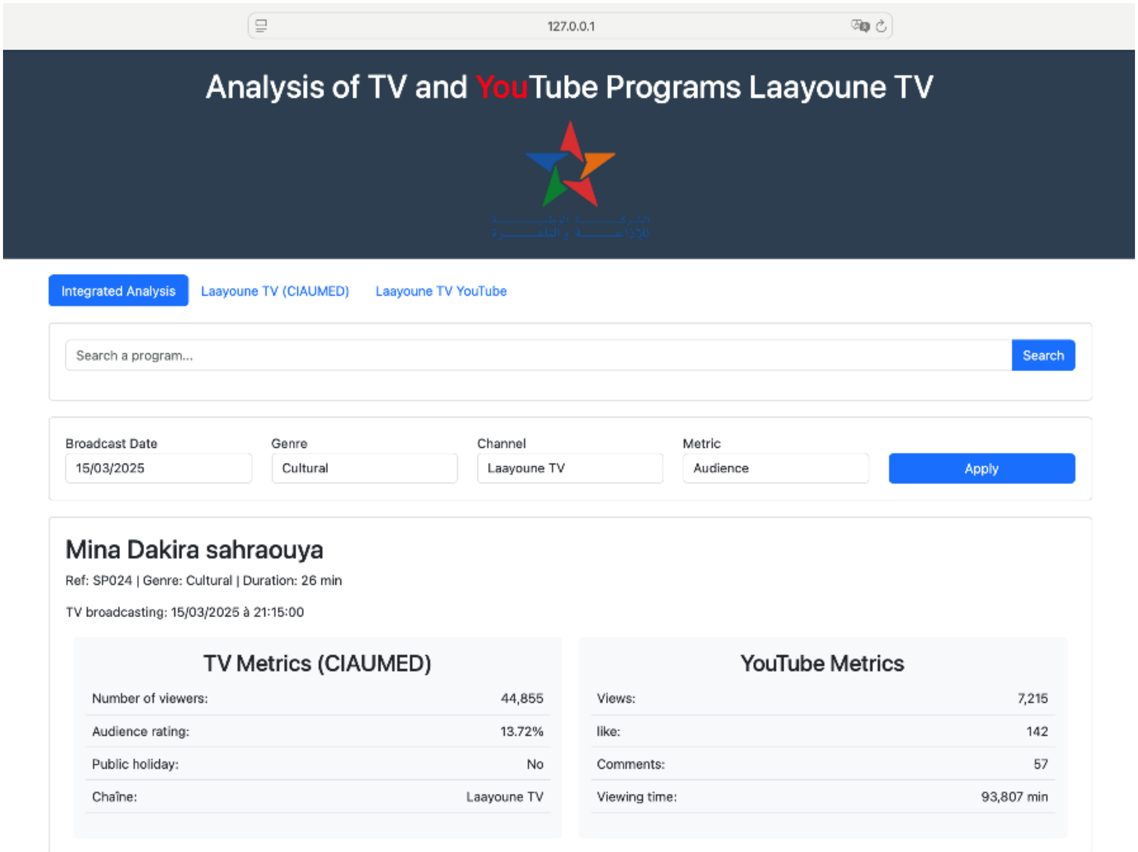

Figure 4. Interface web for predicting TV program success

2.6. Recommendations

The predictive model provides actionable insights that enable TV Laayoune to optimize its programming strategy and enhance audience engagement. By analyzing viewer behavior patterns, the model identifies peak viewing times, allowing the network to strategically schedule popular programs to maximize viewership. Additionally, real-time processing capabilities enable dynamic adjustments to the schedule based on live engagement trends, ensuring the network capitalizes on emerging opportunities. Integrating these insights ensures that high-demand content reaches the largest audience, bolstering overall performance and audience retention.

The integration of traditional and digital metrics offers a deeper understanding of audience preferences, paving the way for personalized content strategies. By tailoring programming blocks to specific viewer demographics and interests, TV Laayoune can increase audience loyalty and satisfaction. Moreover, the model supports targeted advertising by profiling viewer segments, allowing advertisements to be scheduled for maximum relevance and effectiveness. These insights enhance ad performance, boosting revenue generation while ensuring that promotional resources are allocated to programs with the highest growth potential.

Finally, the model's recommendations extend to content creation and promotional campaigns, guiding the development of new programs based on engagement data and identifying underperforming shows that could benefit from targeted marketing efforts. Incorporating direct viewer feedback into the data pipeline ensures continuous refinement of predictions, enabling the network to evolve its strategies and maintain competitiveness. Through these data-driven approaches, can achieve sustained growth, optimize resource allocation, and deliver a more engaging and personalized viewing experience for its audience.

3. RESULTS AND DISCUSSION

The integration of CIAUMED people meter audience metrics with YouTube digital engagement metrics introduces a novel dimension to audience analysis. Traditional metrics provide granular insights into household-level TV viewership, capturing data such as program duration, audience demographics, and channel ratings. In contrast, digital metrics offer a real-time perspective on online audience behavior, including video views, likes, and comments. By combining these structured and unstructured data sources, the integrated dataset provides a holistic view of audience engagement across platforms.

This dual-source approach addresses limitations in prior studies, which often relied solely on either traditional or digital metrics. The ability to bridge in-home TV viewership with online engagement allows the model to capture the hybrid nature of contemporary media consumption, making predictions both more accurate and more actionable. Additionally, this multi-source integration empowers strategic programming by identifying content with crossover appeal and adapting to changing viewing habits.

To ensure an unbiased evaluation, the dataset was divided into training (70%), validation (15%), and test (15%) sets. The model, implemented in TensorFlow, underwent hyperparameter tuning using the validation set. Key parameters-including learning rate, batch size, number of epochs, dropout rate, and kernel size-were optimized to enhance model performance and generalization ability.

3.1. Model performance

The proposed hybrid model achieved robust performance, evidenced by a prediction accuracy of 95%. Key metrics included a MAE of 0.045, RMSE of 0.063, R^2 of 0.89, and MAPE of 4.7%. These results confirm the model's reliability in predicting TV program success, demonstrating its ability to handle complex, integrated data from traditional and digital sources effectively. The performance metrics are detailed in Table 2.

Table 2. Model performance metrics

Metric	Value
MAE	0.045
RMSE	0.063
R^2	0.89
MAPE	4.7%

To ensure robust evaluation, a K-fold cross-validation approach (k=5) was employed. The dataset was divided into five subsets, with each fold serving as the validation set once while the remaining four were used for training. This method minimized bias and variance, providing a thorough assessment of the model's generalizability. The cross-validation process demonstrated consistent performance across folds, with minimal variance observed. For example, MAE values ranged from 0.043 to 0.046, highlighting the model's stability and reliability in real-world applications.

3.2. Comparative analysis

To assess its effectiveness, the hybrid model was compared against baseline models, including linear regression, RF, SVM, KNNs, and a standalone CNN. As shown in Table 3, the hybrid model outperformed all alternatives, demonstrating the benefits of integrating spatial, temporal, and uncertainty modeling techniques. The superior performance of the hybrid model is further illustrated in Figure 5, which visually compares MAE values across different models. By integrating CNNs, LSTMs, and GPs, the proposed model effectively captures spatial, temporal, and uncertainty aspects of audience engagement. Specifically, CNNs extract meaningful patterns from digital engagement data, LSTMs model temporal dependencies in viewership behavior, and GPs provide uncertainty estimation, enhancing the model's interpretability and reliability. This comprehensive approach highlights the importance of multi-source data integration, leading to higher prediction accuracy, improved scheduling strategies, and better decision-making in TV programming.

Table 3. Performance comparison of different models

Model	MAE	RMSE	R^2	MAPE
Linear regression	0.085	0.098	0.72	9.3%
RF	0.067	0.082	0.81	6.8%
SVM	0.073	0.089	0.77	7.5%
KNNs	0.079	0.093	0.74	8.2%
CNN	0.052	0.068	0.85	5.1%
Proposed hybrid model	0.045	0.063	0.89	4.7%

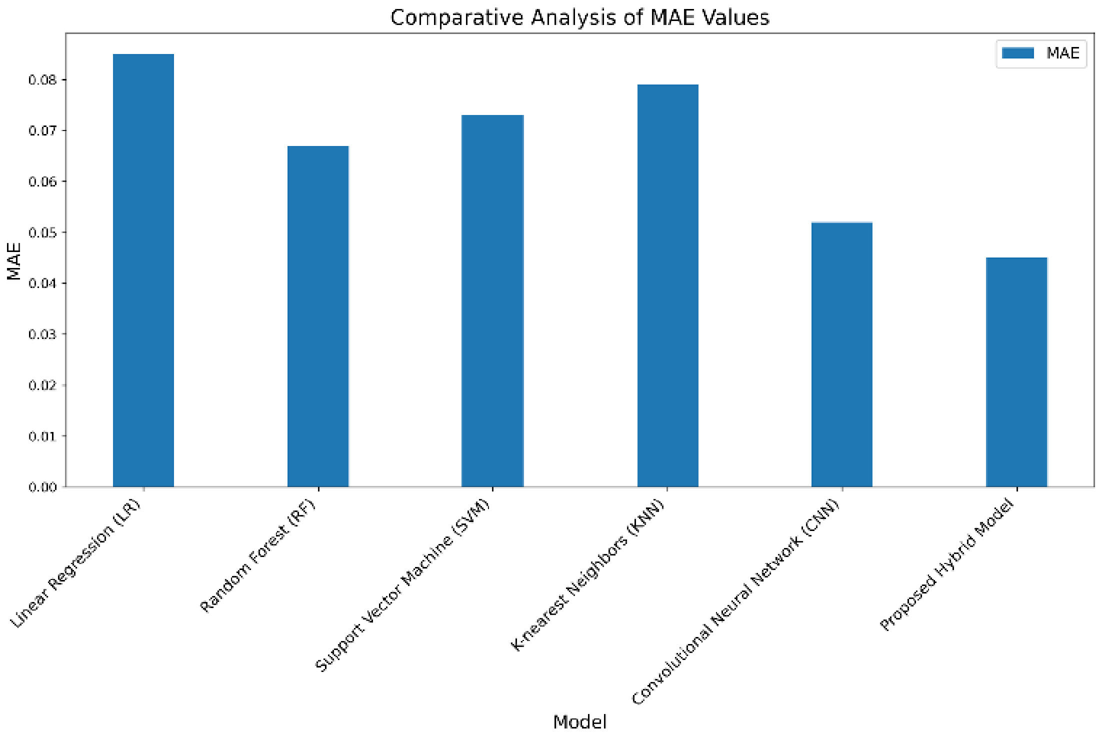

Figure 5. Comparative analysis of MAE values

3.3. Case studies

To further evaluate the practical utility of the proposed model, we conducted an in-depth analysis of three specific TV programs broadcast on Laayoune TV. These case studies demonstrate the model's ability to generate actionable insights that enhance programming strategies and audience engagement. Program A, the model's predictions achieved a 95% alignment with actual audience ratings, confirming its high precision for this program. This demonstrates the model's ability to accurately capture audience preferences for specific content types. Program B, the model correctly forecasted a decline in viewership, which was later confirmed by subsequent audience data. This early warning enabled the network to reallocate marketing and promotional efforts, mitigating the impact of declining audience numbers. Program C, based on the model's predictions, strategic scheduling adjustments were implemented, leading to a 10% increase in viewership. This illustrates how predictive insights can be leveraged to optimize broadcast timing and maximize audience engagement. These examples illustrate the model's ability to provide actionable insights, enhancing programming strategies and optimizing audience engagement.

3.4. Discussion

This study demonstrates the effectiveness of integrating traditional TV audience metrics with digital engagement data to predict program success. Unlike previous models that relied solely on people meter data or digital interactions, our hybrid approach enhances predictive accuracy by combining structured and unstructured data sources. The results confirm that fusing historical TV ratings with real-time online engagement improves forecasting, capturing both long-term audience trends and immediate viewer reactions.

The proposed hybrid model, combining CNNs, LSTMs, and GPs, outperformed all baseline models, achieving an R^2 of 0.89, demonstrating its strong predictive capability. By incorporating CNNs for spatial feature extraction, LSTMs for capturing temporal patterns, and GPs for uncertainty modeling, the model enhances decision-making for TV scheduling, audience targeting, and content strategy. Comparative analysis showed that traditional machine learning models, such as RF and SVM, struggled to capture the dynamic nature of viewership trends, while deep learning models alone, such as standalone CNNs, lacked the ability to model sequential dependencies. The superior performance of the hybrid approach confirms that multi-source integration and deep learning techniques provide a significant advantage in audience prediction.

Despite its strengths, the model has limitations. YouTube engagement data may introduce platform-specific biases, as algorithmic recommendations influence content visibility and interactions. This raises concerns about generalizability, suggesting that future studies should integrate cross-platform engagement data from TikTok, Facebook, and Twitter. Additionally, the computational complexity of the CNN-LSTM-GP model poses challenges for real-time deployment, requiring optimization techniques such as model compression, quantization, and pruning. Finally, since the model was trained on Laayoune TV data, its applicability to other broadcasters and markets remains an open question. Expanding the dataset to global TV networks would improve robustness and adaptability.

Future research could explore social media sentiment analysis and transformer-based architectures (e.g., bidirectional encoder representations from transformers (BERT), generative pre-trained transformer 4 (GPT-4)) to refine audience preference modeling. Additionally, reinforcement learning-based adaptive scheduling could enable dynamic, real-time program adjustments. Addressing these challenges will further enhance artificial intelligence (AI)-driven predictive analytics, optimizing programming strategies and transforming audience engagement in the TV industry.

4. CONCLUSION

This study tackled the challenges of optimizing programming schedules and predicting TV program success for TV Laayoune by integrating people meter audience metrics with digital engagement metrics and employing advanced machine learning techniques. Our findings provide conclusive evidence that combining traditional TV ratings with digital audience interactions enhances forecasting accuracy, bridging the gap between conventional TV consumption and modern online viewing behaviors. By leveraging a hybrid deep learning model integrating CNNs, LSTMs, and GPs, this study demonstrated significant improvements in predicting audience engagement. The results confirmed that multi-source data fusion enhances TV scheduling efficiency and supports data-driven decision-making. This model offers a scalable and adaptable framework for optimizing programming strategies, ensuring content reaches the right audience at the right time.

Beyond improving content scheduling for TV Laayoune, these insights pave the way for more intelligent broadcasting decisions. Future research should focus on expanding dataset diversity across multiple TV networks and incorporating real-time audience interactions to refine adaptability. These advancements will further enhance audience measurement methodologies and support the evolution of AI-driven programming strategies in the TV industry.

ACKNOWLEDGEMENTS

We sincerely thank IBN Tofail University for their continuous support and encouragement. Our gratitude also goes to TV Laayoune for their invaluable resources, collaboration, and practical insights, which were instrumental in developing and implementing our machine learning model for predicting TV program success.

FUNDING INFORMATION

The authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Khalid El Fayq	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓
Said Tkatek		✓		✓	✓	✓			✓	✓	✓	✓	✓	
Lahcen Idouglid				✓		✓		✓		✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY

The data used in this study are confidential and cannot be made publicly available or shared with other parties. These data were used exclusively for the purposes of this research. Due to privacy and confidentiality agreements, access to the dataset is restricted.

REFERENCES

[1] L. Nixon, "Predicting your future audience's popular topics to optimize TV content marketing success," in *AI4TV 2020 - Proceedings of the 2nd International Workshop on AI for Smart TV Content Production, Access and Delivery*, Oct. 2020, pp. 5–10, doi: 10.1145/3422839.3423062.

[2] M. E. Cammarano, A. Guarino, D. Malandrino, and R. Zaccagnino, "TV shows popularity prediction of genre-independent TV series through machine learning-based approaches," *Multimedia Tools and Applications*, vol. 83, no. 31, pp. 75757–75780, Feb. 2024, doi: 10.1007/s11042-024-18518-z.

[3] J. M. Lucas and S. Lazarus, "Audience perception on the impact of television reality shows on the moral lives of adolescents in Taraba State," *International Journal of Humanities, Education, and Social Sciences*, vol. 2, no. 2, pp. 204–230, Jun. 2024, doi: 10.58578/ijhess.v2i2.3242.

[4] K. El Fayq, S. Tkatek, and L. Idouglid, "Predicting television programs success using machine learning techniques," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 5, pp. 5502–5512, Oct. 2024, doi: 10.11591/ijece.v14i5.pp5502-5512.

[5] J. Zhou, Z. Wei, B. Peng, and F. Chi, "Research and application of film and television literature recommendation based on secure internet of things and machine learning," *Mobile Information Systems*, vol. 2021, pp. 1–10, Oct. 2021, doi: 10.1155/2021/4066267.

[6] B. R. Jeyavadhanam, V. V. Ramalingam, V. Sugumaran, and D. Rajkumar, "Predictive analysis of online television videos using machine learning algorithms," in *Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools, and Applications*, Wiley, 2022, pp. 237–257.

[7] G. Verma and H. Verma, "Predicting bollywood movies success using machine learning technique," in *Proceedings - 2019 Amity International Conference on Artificial Intelligence, AICAI 2019*, Feb. 2019, pp. 102–105, doi: 10.1109/AICAI.2019.8701239.

[8] D. O. Oyewola and E. G. Dada, "Machine learning methods for predicting the popularity of movies," *Journal of Artificial Intelligence and Systems*, vol. 4, no. 1, pp. 65–82, 2022, doi: 10.33969/ais.2022040105.

[9] V. Gupta *et al.*, "Predicting attributes based movie success through ensemble machine learning," *Multimedia Tools and Applications*, vol. 82, no. 7, pp. 9597–9626, Mar. 2023, doi: 10.1007/s11042-021-11553-0.

[10] R. A. Abarja, "Movie rating prediction using convolutional neural network based on historical values," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 5, pp. 2156–2164, May 2020, doi: 10.30534/ijeter/2020/109852020.

[11] H. Sharma, S. Pangaonkar, R. Gunjan, and P. Rokade, "Sentimental analysis of movie reviews using machine learning," *ITM Web of Conferences*, vol. 53, p. 02006, Jun. 2023, doi: 10.1051/itmconf/20235302006.

[12] B. Cizmeci and S. G. Oguducu, "Predicting IMDb ratings of pre-release movies with factorization machines using social media," in *UBMK 2018 - 3rd International Conference on Computer Science and Engineering*, Sep. 2018, pp. 173–178, doi: 10.1109/UBMK.2018.8566661.

[13] A. Crisci, V. Grasso, P. Nesi, G. Pantaleo, I. Paoli, and I. Zaza, "Predicting TV programme audience by using Twitter based metrics," *Multimedia Tools and Applications*, vol. 77, no. 10, pp. 12203–12232, May 2018, doi: 10.1007/s11042-017-4880-x.

[14] A. Kupavskii, A. Umnov, G. Gusev, and P. Serdyukov, "Predicting the audience size of a tweet," *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*, vol. 7, no. 1, pp. 693–696, Aug. 2013, doi: 10.1609/icwsm.v7i1.14454.

[15] L. Song, Y. Shi, and G. K. F. Tso, "Commercial audience retention of television programs: measurement and prediction," *International Journal of Advertising*, vol. 41, no. 3, pp. 435–461, Apr. 2022, doi: 10.1080/02650487.2021.1906541.

[16] B. Akgül and T. Küçükyilmaz, “Forecasting TV ratings of Turkish television series using a two-level machinelearning framework,” *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 30, no. 3, pp. 750–766, Mar. 2022, doi: 10.55730/1300-0632.3809.
[17] T. Singh, A. Nayyar, and A. Solanki, “Multilingual opinion mining movie recommendation system using RNN,” in *Lecture Notes in Networks and Systems*, vol. 121, 2020, pp. 589–605.
[18] S. Choi, H. Shin, and S. S. Kang, “Predicting audience-rated news quality: using survey, text mining, and neural network methods,” *Digital Journalism*, vol. 9, no. 1, pp. 84–105, Jan. 2021, doi: 10.1080/21670811.2020.1842777.
[19] K. Lee, J. Park, I. Kim, and Y. Choi, “Predicting movie success with machine learning techniques: ways to improve accuracy,” *Information Systems Frontiers*, vol. 20, no. 3, pp. 577–588, Jun. 2018, doi: 10.1007/s10796-016-9689-z.
[20] N. Ma, S. Zhao, Z. Sun, X. Wu, and Y. Zhai, “An improved ridge regression algorithm and its application in predicting TV ratings,” *Multimedia Tools and Applications*, vol. 78, no. 1, pp. 525–536, Jan. 2019, doi: 10.1007/s11042-017-5250-4.
[21] L. Seric, D. Miletic, A. Ivanda, and M. Braovic, “Predicting TV Viewership with regression models,” in *2022 7th International Conference on Smart and Sustainable Technologies, SpliTech 2022*, Jul. 2022, pp. 1–7, doi: 10.23919/SpliTech55088.2022.9854230.
[22] M. Vijarania, A. Gambhir, D. Sehrawat, and S. Gupta, “Prediction of movie success using sentimental analysis and data mining,” 2022, pp. 174–189.
[23] F. Gegres, D. A. Azar, J. Vybihal, and J. T. L. Wang, “Early prediction of movie success using machine learning and evolutionary computation,” in *2022 21st International Symposium on Communications and Information Technologies, ISCIT 2022*, Sep. 2022, pp. 177–182, doi: 10.1109/ISCIT55906.2022.9931277.
[24] H. Verma and G. Verma, “Prediction model for bollywood movie success: a comparative analysis of performance of supervised machine learning algorithms,” *The Review of Socionetwork Strategies*, vol. 14, no. 1, pp. 1–17, Apr. 2020, doi: 10.1007/s12626-019-00040-6.
[25] R. Akula, Z. Wieselthier, L. Martin, and I. Garibay, “Forecasting the success of television series using machine learning,” in *Conference Proceedings - IEEE SOUTHEASTCON*, Apr. 2019, vol. 2019-April, pp. 1–8, doi: 10.1109/SoutheastCon42311.2019.9020419.
[26] K. El Fayq, S. Tkatek, and L. Idouglid, “Towards emotional authenticity in news presentation: a machine learning approach,” in *2024 11th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Jul. 2024, pp. 1–7, doi: 10.1109/wincom62286.2024.10654995.

BIOGRAPHIES OF AUTHORS

Khalid El Fayq was born in Laayoune, Morocco, in 1985. He obtained his engineering degree in computer science in 2020 from Agadir International University. Currently, he is a Ph.D. student at the Computer Engineering Training Department (LaRIT), Faculty of Science, Ibn Tofail University, Kenitra, Morocco. He works in the media field and serves as a temporary teacher at universities. His research interests encompass artificial intelligence and the audiovisual field. He can be contacted at email: khalidelfayq@gmail.com.

Said Tkatek is a Professor of Computing Science Research at the Faculty of Science, Ibn Tofail University, Kenitra, Morocco. He is a member of the Computer Science Research Laboratory (LaRIT). His research focuses on the optimization of NP-hard problems using metaheuristic approaches in artificial intelligence and big data for decision-making across various fields. This includes big data analytics, artificial intelligence, and information systems. He can be contacted at email: said.tkatek@uit.ac.ma.

Lahcen Idouglid a Ph.D. researcher at Ibn Tofail University’s Faculty of Sciences in Kenitra, works within the Computer Science Research Laboratory (LaRIT). His research encompasses computer networks, software engineering, artificial intelligence, and security. He can be contacted at email: lahcen.idouglid@uit.ac.ma.