Predicting television programs success using machine learning techniques

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

Article Info

Article history:

Received Aug 15, 2023
Revised Jun 17, 2024
Accepted Jul 2, 2024

Keywords:

Data
Data science
Forecast models
Machine learning
Predicting
Television
Television program

ABSTRACT

In the ever-evolving media landscape, television (TV) remains a coveted platform, compelling industry players to innovate amid intense competition. This study focuses on leveraging machine learning regression models to precisely predict TV program reach. Our objective is to assess the models' efficacy, revealing a standout performer with a mean absolute percent error of just under 8%. Significantly, we identify features exerting a substantial impact on predictions and explore the potential for model enhancement through expanded datasets. This research extends beyond statistical insights, offering actionable implications for TV channel managers. Empowered by these findings, managers can make informed decisions in program planning and scheduling, optimizing viewer engagement. The temporal analysis of evolving trends over time adds a nuanced layer to our study, aligning it with the dynamic nature of the media landscape. As television retains its dynamic force, our insights contribute not only to academic discourse but also provide practical guidance, enhancing the competitive edge of television channels.

Corresponding Author:

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

1. INTRODUCTION

The marketing industry stands as one of the largest in the world. Television (TV) companies invest millions of dollars in marketing, and audience ratings can assist advertisers in obtaining the information they need to create ads that meet their audience's needs. This information also provides content creators and TV networks with the ability to make improvements and plays a crucial role in various areas, including content promotion and programming.

In Morocco, the audience measurement system stands at the forefront of innovation. The current audience measurement methods in Morocco operate on the principle of sampling. Several key organizations play pivotal roles in this process, including the High Authority for Audiovisual Communication (HACA), which is the regulatory body overseeing audiovisual communication, the Interprofessional Media Audience Center (CIAUMED), responsible for establishing and implementing the media measurement system for audiences in Morocco. Additionally, Marocmétrie is entrusted with the crucial responsibility of collecting and processing media audience data in Morocco, and the national radio and television company (SNRT) is actively engaged, overseeing the measurement of audiences for all Moroccan channels across the entire Moroccan territory, encompassing both urban and rural demographics.

To contextualize this study within the broader field of television programming and audience measurement, it is imperative to acknowledge the state-of-the-art. Existing studies, exemplified by [1]–[7], predominantly focus on predicting impressions for specific content, movies, series, and TV ratings. However, a discernible gap emerges in the arena of predicting the overall reach of a diverse audience. This paper endeavors to address this gap by employing predictive analytics, specifically aiming to enhance the overall TV audience percentage. Drawing insights from a specialized company in Moroccan television audience statistics, we seek to contribute novel perspectives to the existing body of knowledge.

Laayoune channel, one of the Moroccan television channels, has its daily TV guides manually scheduled by the staff within the prediction department. a process that can be tedious and prone to human error. At times, shows are scheduled incorrectly, creating inaccuracies in the TV schedule. This may result from misjudgments of demographic data or TV ratings. Such errors can lead, for example, to incorrect pricing for advertisements or contract breaches, resulting in financial losses and decreased viewership.

The following outlines relevant works in the field of machine learning solutions for prediction and classification within the domain of audiovisual content [8], [9], Accurately forecasting audience ratings is paramount for determining a TV show's revenue, offering benefits like efficient investment planning. Despite the evident need for precise TV prediction [10], [11], this area remains relatively underexplored in practice. Recent advancements in data mining and machine learning have spurred studies to apply novel techniques to the TV ratings prediction problem. Notably, since 2000, there has been a focus on individual rating behavior, including an analysis of social media platforms such as Facebook and Twitter, incorporating predictors specific to TV programs or series, such as post counts, likes, shares, and comments for specified pages [12]–[16].

In the realm of predicting TV ratings, studies utilize aggregated data for individuals [17], [18], and people meter values for all households [19]. Various techniques are applied to address the TV rating prediction problem, including neural networks [20]–[22], decision trees [23], [24], and gradient-boosting machines [25]. Ridge regression [26], genetic algorithm [27], [28], and linear regression [29], are frequently employed. Another study on machine learning prediction examined the possibility of predicting the popularity of new movies and series [30]–[36]. In this study, we tested several different prediction models such as random forest (RF), k-nearest neighbors (KNN), support vector machine (SVM), and linear regression (LR).

This paper introduces a paradigm-shifting forecasting module tailored for the Laayoune channel. At its core lies the integration of supervised machine learning, a powerful tool to automate the identification of optimal broadcast times. Trained meticulously on historical data, the machine learning model embarks on an insightful analysis of past broadcasting habits, identifies recurring patterns, and anticipates the ideal broadcast times for specific shows. Our study makes a substantial contribution to the field by addressing the challenges faced by the Laayoune channel, including manual scheduling errors, financial losses, and decreased viewership. The introduction of this automated program represents a significant leap forward, not only reducing the labor-intensive nature of creating the TV guide but also elevating accuracy and minimizing potential errors associated with manual planning.

The remainder of this paper is organized as follows: section 2 describes the proposed approach, including problem formulation, data description, and TV broadcast preparation. Section 3 is dedicated to the presentation and analysis of results, offering a nuanced understanding of predictive models for top-rated TV shows. Section 4 concludes with insights into future extensions and applications of our methodology.

2. METHOD

This chapter will present the methods and the arguments employed in this study along with the justifications for their use in this paper. A comprehensive literature review was undertaken to identify and assess various types of machine learning models applicable to the given problem. This involved an examination of prior studies, research, and work related to supervised machine learning. We then assessed and compared the chosen models using diverse methods. Finally, we collected and analyzed the results of the evaluation and comparison.

The choice to employ a supervised machine learning model was grounded in its capacity to predict data based on its prior knowledge of the dataset. Extensive research and analysis of various predictive models revealed that supervised learning methods were more suitable for the task at hand compared to other approaches. Furthermore, the nature of the data used in similar studies aligned with the data available in this project, reinforcing the preference for supervised models.

Based on the research reviewed, we decided to utilize k-nearest neighbors regression, decision tree, random forest, linear regression, support vector machine, multi-layer perceptron (MLP) neural network, and ridge regression. These models were deemed most appropriate based on their performance and applicability in previous studies. By employing a combination of these chosen models, we aim to achieve accurate predictions for the number of viewers and audience share, facilitating informed decision-making for television programs.

The methodology put forward involves six specific phases: data preprocessing, extracting features, constructing the predictive model, assessing the model's performance, and implementing the model. As shown in Figure 1, the overall workflow of the process. Initially, the raw data undergoes preprocessing. This is followed by feature extraction to facilitate the process. Subsequently, the dataset is divided into training and test sets. The machine learning models are then trained using the training data. After training, each model's performance is assessed through testing. Finally, the model exhibiting high performance is deployed using the Flask framework for user interaction.

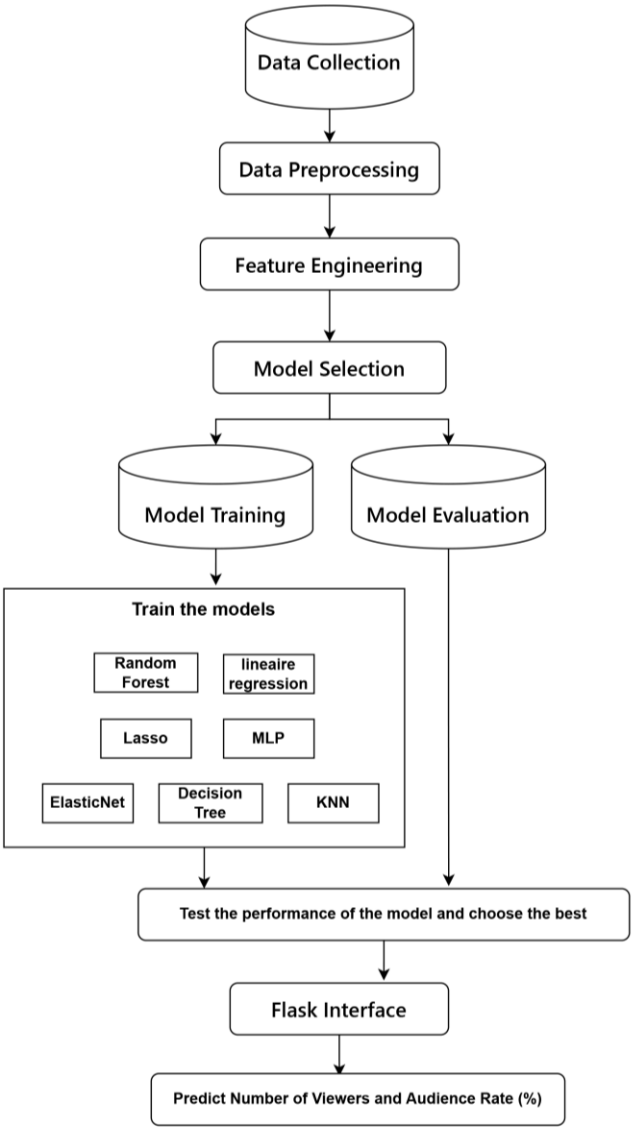


Figure 1. Proposed model

2.1. Getting data raw

Given the impracticality of directly measuring the true television viewing rate, the utilized TV rating values are those published by CIAUMED. The acquired dataset comprises historical data from the year 2020, featuring the attributes (columns) illustrated in Figure 2. Furthermore, the dataset is categorized into twelve distinct classes, as depicted in Figure 3.

The audience measurement organization in Morocco uses a statistical method called "Mediametrie" to assess television channel viewership. This technique relies on a carefully selected panel of individuals representing the country's television audience. These panelists are equipped with measurement devices, such as boxes or software, which automatically record the programs they watch on television. A common practice in audience measurement is the use of inaudible audio codes, also known as audio identification codes (AIC) or audio watermarking codes. Inserted into the broadcast signal of television channels which are shown in

Figure 4 from the control room of SNRT, these codes are designed to be inaudible to viewers, ensuring an undisturbed viewing experience.

AIC plays a crucial role in médiamétrie by providing a method to mark audio content. When a program is broadcast, the corresponding audio identification code is also transmitted with the content. Measurement devices with viewers are designed to detect these inaudible codes. When a code is detected, information about the watched program is recorded and transmitted to the measurement organization. Thus, the data collected from the representative panel allows CIAUMED to estimate the total number of viewers watching each channel at any given time, along with their demographic profile.

	Program Ref	Program Name	Genre Program	Number of viewers	Audience share (%)	Start time	Broadcasting date	Duration	Public_Holiday
0	5	Assrisser Dahbo	Series	205665	41	19:15:00	2020-01-01	26	0
1	10	Journal Télévisé	News	39855	11	22:30:00	2020-01-01	30	0
2	12	Konoz mina sahrae	Documentaries	36491	19	10:00:00	2020-01-01	26	0
3	20	Makam mina tarab hassani	Music	146069	44	19:40:00	2020-01-01	26	0
4	29	Rab Ojab	Cultural	47455	31	21:10:00	2020-01-01	26	0

Figure 2. Example of data from CIAUMED

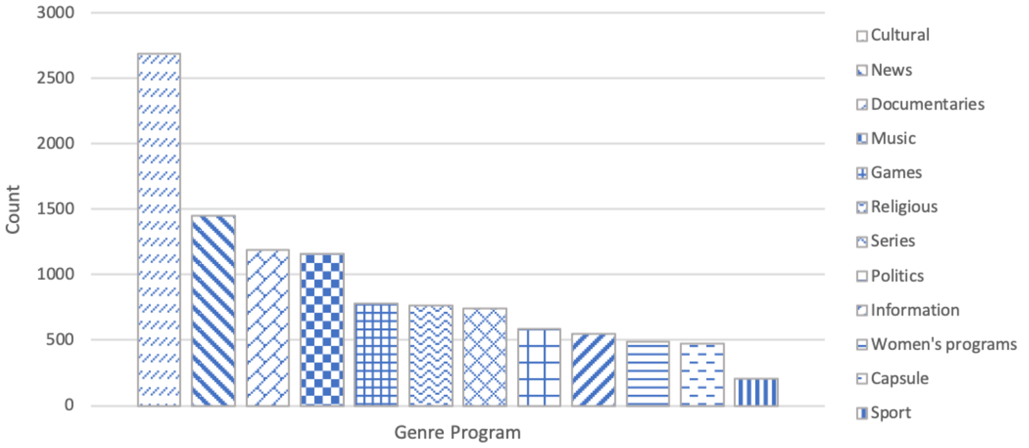


Figure 3. Genre distribution of TV programs

Figure 4. Audience measurement equipment in the control room (SNRT)

The generation of audio identification codes uses various techniques, including phase modulation. This method involves subtle variations in the phase of the original audio signal, making the codes inaudible to human listeners but detectable by measurement or tracking devices. Other approaches, such as Fourier transformation, frequency modulation, and pseudo-random codes, are also employed to create AIC. These audio identification codes are designed to be robust against common audio distortions and processing, ensuring accurate detection even in cases of altered transmission or reproduction of the audio signal.

2.2. Data features and formatting

Before building a score prediction model, data preprocessing is a critical and fundamental step in the data analysis and machine learning pipeline. It involves transforming raw data into a clean, organized, and structured format that is suitable for analysis and modeling. During this stage, various techniques are applied to handle missing values, outliers, and noise, ensuring data quality and reliability. Additionally, data normalization and scaling may be performed to bring all features to a common scale, preventing any feature from dominating the others. Categorical variables might be encoded into numeric representations, enabling algorithms to process them effectively. Data preprocessing sets the stage for accurate feature extraction and model building, enhancing the overall performance and interpretability of the predictive models. The acquired data set contains many variables, and in our study, we need:

- Program Ref: this code is used to uniquely identify and reference each program.
- Program: The name of the TV program
- Genre: Genre of the TV show
- Number of viewers: The number of viewers for each TV program
- Audience rate (%): Viewership rating compared to other TV programs on the other Moroccan TV station
- Start time: The start time of the broadcast
- Broadcast date: The date of program broadcasting
- Duration: The duration of the TV program in minutes
- Public_Holiday: This column is used to mark if a TV program was broadcast on a public holiday.

The datasets illustrate the start time and date of a program within a day, as shown in Figure 5(a) and 5(b). These figures show two parts of the same table, split for clarity. Changes to the data involved eliminating “:” to avoid confusion with "/", facilitating the transformation of text into numerical values. This method was similarly employed in categorizing the dataset based on TV program categories.

	Program Name	Genre Program	Number of viewers	Audience share (%)	Duration	Program Name Frequency	Genre Program Frequency	Program Ref Frequency	Broadcasting year	Broadcasting month	...	Actual Number of Viewers	Time Slot_Evening
0	Agenda	Information	0.113163	0.12500	0.062500	0.491597	0.185078	0.491597	-12.464336	0.845580	...	0.014145	0
1	Fada Ryadi	Sport	0.137256	0.03125	1.000000	0.472689	0.053295	0.472689	-2.145805	-1.521047	...	0.004289	0
2	Journal Télévisé	News	0.232234	0.09375	0.479167	1.000000	0.296512	1.000000	-2.145805	-1.521047	...	0.021772	1
3	Makam mina tarab hassani	Music	0.421529	0.40625	0.395833	0.325630	0.350775	0.325630	-2.145805	-1.521047	...	0.171246	0
4	Assrisser Dahbo	Series	0.305037	0.31250	0.395833	0.294118	0.066860	0.294118	-2.145805	-1.521047	...	0.095324	0

(a)

Program Ref Frequency	Broadcasting year	Broadcasting month	...	Actual Number of Viewers	Time Slot_Evening	Time Slot_Late Night	Time Slot_Morning	Day of the Week_1	Day of the Week_2	Day of the Week_3	Day of the Week_4	Day of the Week_5	Day of the Week_6
0.491597	-12.464336	0.845580	...	0.014145	0	0	0	0	0	1	0	0	0
0.472689	-2.145805	-1.521047	...	0.004289	0	0	0	0	0	1	0	0	0
1.000000	-2.145805	-1.521047	...	0.021772	1	0	0	0	0	1	0	0	0
0.325630	-2.145805	-1.521047	...	0.171246	0	1	0	0	0	0	0	0	1
0.294118	-2.145805	-1.521047	...	0.095324	0	0	1	0	0	0	0	0	0

(b)

Figure 5. Data used for prediction models (a) first half and (b) second half

2.3. Feature engineering

After the data undergoes cleaning and tokenization, it becomes crucial to extract features from this processed data. This is because machines are incapable of comprehending words in their raw form and can only interpret numerical values. Feature engineering stands as a vital step in the data preprocessing sequence, focusing on the generation of new features or the modification of current ones. This step is essential for improving the efficacy of machine learning algorithms. This process involves selecting, transforming, and creating features from the raw data that best capture the underlying patterns and relationships. By deriving meaningful features, the predictive power of the models can be significantly improved, leading to better insights and accurate predictions. It requires domain knowledge and a deep understanding of the data to intelligently engineer features that reflect the nuances and intricacies of the problem domain.

2.4. Training process

After checking the distribution of data in Figure 6, each element with its frequency, a training split of 80/20 was used in the training process. 80% of the data in the dataset was used for training and 20% of the data in the dataset was used for testing after training the model. The decision to adopt the 80/20 split was based on the superior results it demonstrated compared to other tested splits. To ensure the robustness of our approach, we conducted multiple training iterations for each model, systematically examining if there were any variations in prediction outcomes.

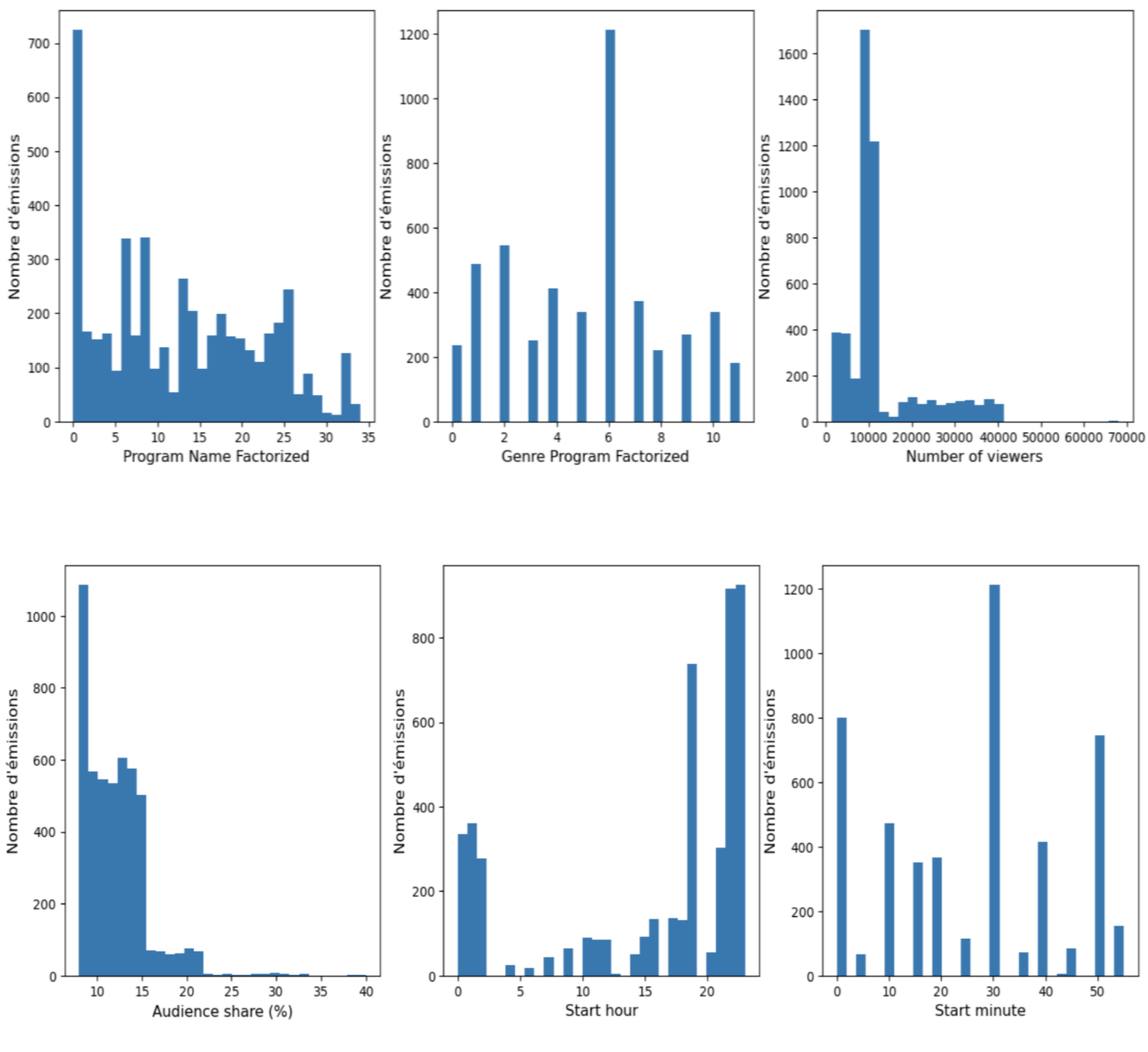


Figure 6. Plot visualization for checking the distribution of data

2.5. Testing process

The testing process primarily focused on prediction accuracy to assess the relative performance of the models. The primary objective was to examine how well the models compared in terms of prediction accuracy. The model demonstrating the best predictive accuracy was carefully selected for further analysis. The model selection criterion was based on predictive accuracy, and the best-performing model among others was evaluated using the mean squared error (MSE). The use of MSE allows for a quantitative and robust evaluation of model performance, taking into account the accuracy of predictions across the entire dataset. The ultimate goal of this model selection process is to find a model that excels not only on training data but also generalizes effectively to make accurate predictions on real-world data. The key steps in the model selection process in this project can be summarized as follows:

2.5.1. Model evaluation

To commence the model evaluation process, we start by considering a variety of candidate algorithms. These include linear regression, random forests, decision trees, Ridge, ElasticNet, k-nearest neighbors, and neural networks. It is crucial to note that each algorithm comes with its own set of strengths and weaknesses, making it imperative to comprehensively test multiple options.

2.5.2. Hyperparameter tuning and cross-validation

In the context of the television program prediction project, an advanced and robust evaluation strategy is implemented by combining hyperparameter tuning with cross-validation. This dual approach ensures a thorough exploration of model performance. The crucial steps involved in hyperparameter tuning enable the identification of optimal model configurations, while cross-validation enhances the reliability of the results by testing the models across various subsets of the dataset. Through these carefully orchestrated steps, we aim to pinpoint the most accurate and reliable predictive models, a critical factor contributing to the overall success of the project.

2.5.3. Comparison and selection

By combining hyperparameter tuning with cross-validation, a more comprehensive and rigorous approach is achieved to evaluate the performance of machine learning models in the television program prediction project. These steps help identify the best model configurations and hyperparameters to obtain the most accurate and reliable predictions, contributing to the overall success of the project. After using appropriate performance metrics for prediction tasks, such as mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE), we compared the performance of different models based on the chosen metrics. In our analysis, we carefully considered the trade-off between bias and variance, aiming to select a model that minimizes overfitting while capturing underlying patterns. Once the optimal model was identified, it was trained on the entire dataset, and predictions were made on new data to validate its effectiveness.

2.6. Tools and framework

In the context of a prediction project, a diverse set of tools and frameworks is commonly employed to execute essential tasks, including data manipulation, the development of predictive models, and the evaluation of results. Some popular tools and frameworks include scikit-learn, Pandas, NumPy, and data visualization libraries like Matplotlib and Seaborn. Scikit-learn is an open-source library that offers tools for data analysis, prediction, and machine-learning models. Pandas, another open-source software library, provides a set of tools for data analysis and manipulation. This library facilitates data manipulation, making it suitable for preparing data for learning in Scikit-learn's predictive models. NumPy, on the other hand, is a framework for scientific computing in Python. It extends support for large, multi-dimensional arrays and matrices, along with a comprehensive collection of high-level mathematical functions for array operations. NumPy plays a crucial role in converting ".csv" data into array objects, essential for generating training and test datasets for the machine learning models.

3. RESULTS AND DISCUSSION

This chapter delineates the results evaluation process. We conducted a comparative analysis of the models to identify the most effective one. Subsequently, we juxtaposed the top-performing models with expert evaluations to assess the accuracy of TV predictions. In the course of the experiments, we initially tested all the trained models on test data. Following the identification of the best model, we deployed it using the essay grading web application, constructed using Python Flask. As shown in Figure 7 our model was encapsulated in a pickle file, facilitating seamless utilization by the prediction team at the television channel.

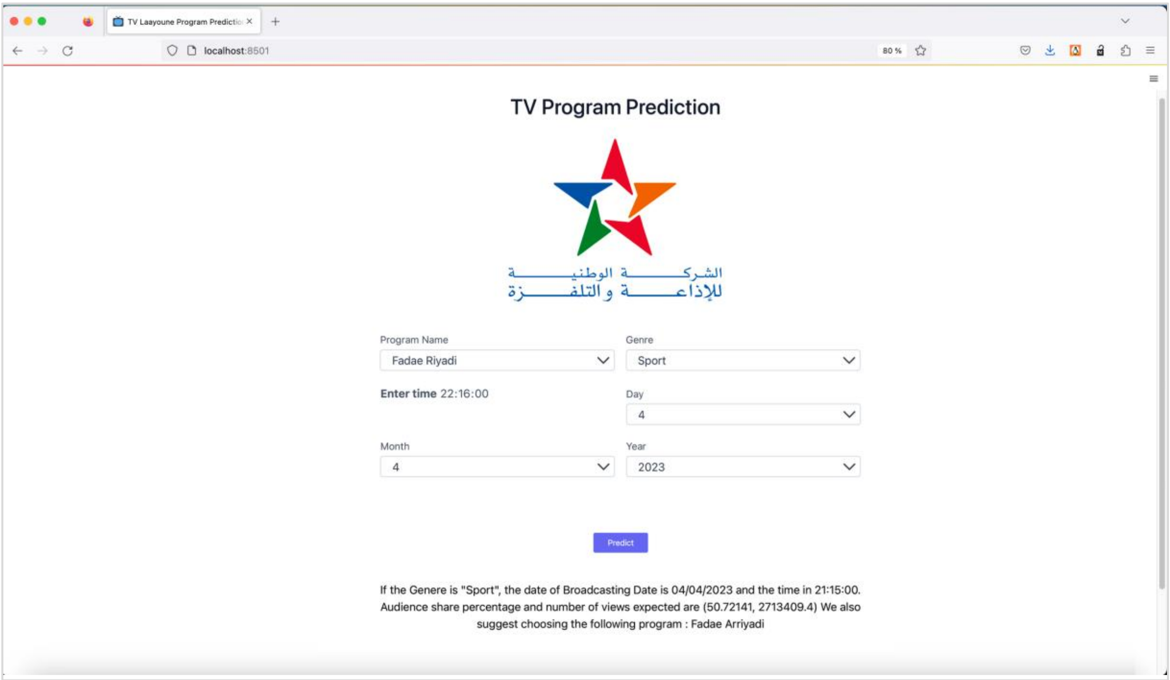


Figure 7. Extract result in a system report

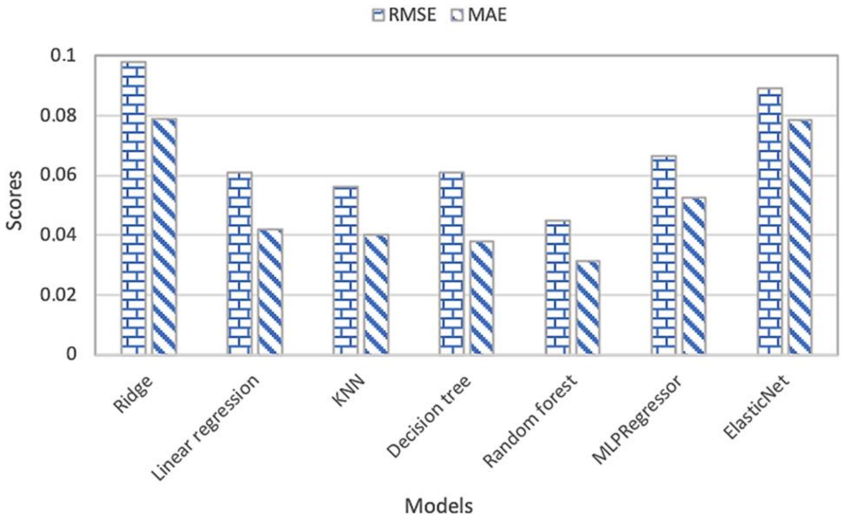
Table 1 provides a comprehensive overview of the performance scores for various algorithms tested in the study. Notably, random forest and k-nearest neighbors emerge as top performers, exhibiting the highest performance scores among the tested algorithms. Their superior performance, as indicated in the table, underscores their effectiveness when compared to other algorithms.

Table 1. Score for the best algorithm performance

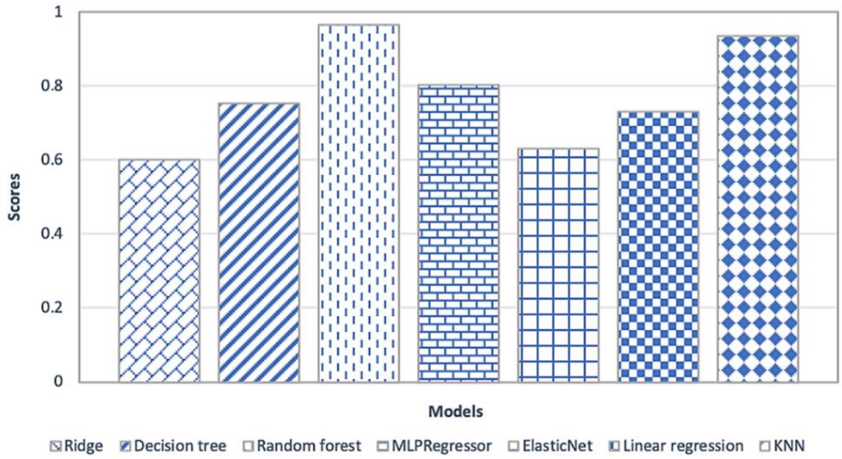
Model	RMSE	MAE	R2 Score
Ridge	0.098081	0.078808	0.601777
Linear regression	0.061159	0.041841	0.731777
KNN	0.056330	0.039984	0.935213
Decision tree	0.060985	0.037925	0.752115
Random forest	0.044828	0.031368	0.966380
MLPRegressor	0.066397	0.052581	0.803228
ElasticNet	0.089067	0.078565	0.631261

In evaluating machine learning algorithms for predicting television programs' success, our training and testing process included using pseudorandom subsets from the dataset to prevent overfitting and ensure the models' generalization capabilities. The random forest regressor algorithm demonstrated superior performance on the test set with a RMSE of 0.09653, MAE of 0.056342, and an R2-score of 0.8980, indicating a high level of precision in its predictions. Notably, when the random forest regressor was assessed across the entire dataset, it continued to show exceptional performance, achieving an RMSE of 0.044828, which suggests consistent and reliable predictions. The R2 score of 0.9663 also indicates that the model explains a significant proportion of the variance in the target variable, emphasizing its effectiveness and robust predictive capability.

To visually elucidate our analytical outcomes, Figure 8 is segmented into two subfigures; Figure 8(a) showcases a comparative assessment of the algorithms based on RMSE and MAE, while Figure 8(b) delineates the comparison based on R2 scores, clearly highlighting the RF model's superior performance across both metrics. This observation initially suggests the efficacy of RF in predicting television program success. However, for a comprehensive evaluation, we also consider the KNN model for a detailed comparison against existing methods, focusing on optimizing TV show scheduling strategies based on date and time to maximize viewership. This approach underscores our commitment to leveraging data-driven insights for enhancing the strategic planning of television programming.



(a)



(b)

Figure 8. Comparative analysis of predictive models in (a) RMSE and MAE across different algorithms and (b) R2 score comparison among algorithms

4. CONCLUSION

This study delved into the potential of leveraging supervised machine learning to predict optimal TV shows for broadcasting, enhance viewership, and optimize the TV Guide schedule. Evaluation of various supervised machine learning models revealed that random forest and k-nearest neighbors regression exhibited the most effective performance, yielding an impressive 97% accuracy in predicting the success of TV programs at a given moment. After performing regression analysis, our model demonstrated superior predictive capabilities, particularly when trained at 80% and tested at 20%, resulting in the least MAE, RMSE, and MSE values for a majority of TV programs.

The study utilized a dataset spanning three years, comprising just under 12,500 rows of data. The machine learning model's performance might have improved with a larger dataset for training and testing. Although the model in this paper does not replace the current process, it could serve as a valuable tool for obtaining quick predictions.

While the results are promising, albeit not surpassing the accuracy of the current TV ratings estimation process, there is potential for machine learning models to replace the existing procedures in the future. Future research in this area should involve employing significantly larger datasets to assess their impact on model performance. Additionally, exploring a broader range of machine learning models and integrating the study dataset with internet TV streaming sources such as YouTube and the channel's website could provide insights into further improving performance.

In future endeavors, we plan to introduce enhancements through the adoption of more advanced techniques, specifically incorporating hybridization with genetic algorithms. This strategic approach holds the promise of achieving finer optimization of our models by dynamically adjusting parameters. This, in turn, will pave the way for even more accurate predictions that are finely tuned to the nuanced dynamics of the ever-evolving television landscape.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude for the supervision and resources provided by esteemed institutions, including the General Directorate of the National Broadcasting and Television Company of Morocco - SNRT (*http://www.snrt.ma*) and the Faculty of Sciences at Ibn Tofail University - FS UIT (*https://fs.uit.ac.ma*).

REFERENCES

[1] 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, Aug. 2016, doi: 10.1007/s10796-016-9689-z.
[2] J. B. Rebecca, V. V Ramalingam, V. Sugumaran, and D. Rajkumar, "Predictive analysis of online television videos using machine learning algorithms," *Fundamentals and Methods of Machine and Deep Learning*. Wiley, pp. 237–257, Jan. 2022. doi: 10.1002/9781119821908.ch10.
[3] Y. An, J. An, and S. Cho, "Artificial intelligence-based predictions of movie audiences on opening Saturday," *International Journal of Forecasting*, vol. 37, no. 1, pp. 274–288, Jan. 2021, doi: 10.1016/j.ijforecast.2020.05.005.
[4] 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.
[5] R. Akula, Z. Wieselthier, L. Martin, and I. Garibay, "Forecasting the success of television series using machine learning," *2019 SoutheastCon*. IEEE, Apr. 2019. doi: 10.1109/southeastcon42311.2019.9020419.
[6] A. Arslan *et al.*, "A comprehensive framework for measuring the immediate impact of TV advertisements: TV-Impact," *Entropy*, vol. 26, no. 2, p. 109, Jan. 2024, doi: 10.3390/e26020109.
[7] 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*, Feb. 2024, doi: 10.1007/s11042-024-18518-z.
[8] E. C. Alemán Carreón, H. Nonaka, A. Hentona, and H. Yamashiro, "Measuring the influence of mere exposure effect of TV commercial adverts on purchase behavior based on machine learning prediction models," *Information Processing & Management*, vol. 56, no. 4, pp. 1339–1355, Jul. 2019, doi: 10.1016/j.ipm.2019.03.007.
[9] C. E. G. Moreta, M. R. C. Acosta, and I. Koo, "Prediction of digital terrestrial television coverage using machine learning regression," *IEEE Transactions on Broadcasting*, vol. 65, no. 4, pp. 702–712, Dec. 2019, doi: 10.1109/tbc.2019.2901409.
[10] L. Wang, "Forecast model of TV show rating based on convolutional neural network," *Complexity*, vol. 2021, pp. 1–10, Feb. 2021, doi: 10.1155/2021/6694538.
[11] P. Danaher and T. Dagger, "Using a nested logit model to forecast television ratings," *International Journal of Forecasting*, vol. 28, no. 3, pp. 607–622, Jul. 2012, doi: 10.1016/j.ijforecast.2012.02.008.
[12] Y.-H. Cheng, C.-M. Wu, T. Ku, and G.-D. Chen, "A predicting model of TV audience rating based on the Facebook," *2013 International Conference on Social Computing*. IEEE, Sep. 2013. doi: 10.1109/socialcom.2013.167.
[13] T. Thaipisutikul and S. Tuarob, "Beyond the tweets: Discovering factors that influence TV series preferences from ubiquitous social networks," *2017 10th International Conference on Ubi-media Computing and Workshops (Ubi-Media)*. IEEE, Aug. 2017. doi: 10.1109/umedia.2017.8074106.
[14] C. Akarsu and B. Diri, "Turkish TV rating prediction with Twitter," *2016 24th Signal Processing and Communication Application Conference (SIU)*. IEEE, May 2016. doi: 10.1109/siu.2016.7495748.
[15] B. Cizmeci and S. G. Oguducu, "Predicting IMDb Ratings of Pre-release Movies with factorization machines using social media," *2018 3rd International Conference on Computer Science and Engineering (UBMK)*. IEEE, 2018. doi: 10.1109/ubmk.2018.8566661.
[16] M.-H. Cheng, Y.-C. Wu, and M.-C. Chen, "Television meets Facebook: The correlation between TV ratings and social media," *American Journal of Industrial and Business Management*, vol. 06, no. 03, pp. 282–290, 2016, doi: 10.4236/ajibm.2016.63026.
[17] 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. 2021, doi: 10.1080/02650487.2021.1906541.
[18] M. J. Awan *et al.*, "A recommendation engine for predicting movie ratings using a big data approach," *Electronics*, vol. 10, no. 10, p. 1215, May 2021, doi: 10.3390/electronics10101215.
[19] L. Nixon, "Predicting your future audience's popular topics to optimize TV content marketing success," *Proceedings of the 2nd International Workshop on AI for Smart TV Content Production, Access and Delivery*. ACM, Oct. 2020. doi: 10.1145/3422839.3423062.
[20] S. Tkatek, S. Bahti, Y. Lmzouari, and J. Abouchabaka, "Artificial intelligence for improving the optimization of np-hard problems: A review," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 5, pp. 7411–7420, Oct. 2020, doi: 10.30534/ijatcse/2020/73952020.
[21] 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.
[22] P. Zhang, C. Liu, K. Ning, W. Zhu, and Y. Zhang, "Prediction and profiling of audience competition for online television series," *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, Aug. 2020. doi: 10.1145/3394486.3403188.
[23] L. Song, Y. Shi, G. K. F. Tso, and H. P. Lo, "Forecasting week-to-week television ratings using reduced-form and structural dynamic models," *International Journal of Forecasting*, vol. 37, no. 1, pp. 302–321, Jan. 2021, doi: 10.1016/j.ijforecast.2020.06.002.
[24] K. Pradeep, C. R. TintuRosmin, S. S. Durom, and G. S. Anisha, "Decision tree algorithms for accurate prediction of movie

rating," *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, Mar. 2020. doi: 10.1109/iccmc48092.2020.iccmc-000158.

[25] S. M. R. Abidi, Y. Xu, J. Ni, X. Wang, and W. Zhang, "Popularity prediction of movies: from statistical modeling to machine learning techniques," *Multimedia Tools and Applications*, vol. 79, no. 47–48, pp. 35583–35617, Jan. 2020, doi: 10.1007/s11042-019-08546-5.

[26] Q. Chen and S. Lee, "A machine learning approach to predict customer usage of a home workout platform," *Applied Sciences*, vol. 11, no. 21, p. 9927, Oct. 2021, doi: 10.3390/app11219927.

[27] H.-L. Huang *et al.*, "Predicting television ratings and its application to Taiwan cable TV channels," *Advances in Intelligent Systems Research*. Atlantis Press, 2013. doi: 10.2991/3ca-13.2013.48.

[28] S. Tkatek, A. Otman, J. Abouchabaka, and N. Rafalia, "A hybrid heuristic method to solve an assignment problem of human resource," *International Review on Computers and Software (IRECOS)*, vol. 10, no. 9, p. 977, Nov. 2015, doi: 10.15866/irecos.v10i9.7246.

[29] L. Seric, D. Miletic, A. Ivanda, and M. Braovic, "Predicting TV viewership with regression models," *2022 7th International Conference on Smart and Sustainable Technologies (SpliTech)*. IEEE, Jul. 2022. doi: 10.23919/splitech55088.2022.9854230.

[30] G. Verma and H. Verma, "Predicting bollywood movies success using machine learning technique," *2019 Amity International Conference on Artificial Intelligence (AICAI)*. IEEE, Feb. 2019. doi: 10.1109/aicai.2019.8701239.

[31] S. Mundra, A. Dhingra, A. Kapur, and D. Joshi, "Prediction of a movie's success using data mining techniques," in *Smart Innovation, Systems and Technologies*, Springer Singapore, 2018, pp. 219–227. doi: 10.1007/978-981-13-1742-2_22.

[32] T. Kim, J. Hong, and P. Kang, "Box office forecasting using machine learning algorithms based on SNS data," *International Journal of Forecasting*, vol. 31, no. 2, pp. 364–390, Apr. 2015, doi: 10.1016/j.ijforecast.2014.05.006.

[33] 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.

[34] H. Sharma, S. Pangaonkar, R. Gunjan, and P. Rokade, "Sentimental analysis of movie reviews using machine learning," *ITM Web of Conferences*, vol. 53, p. 2006, 2023, doi: 10.1051/itmconf/20235302006.

[35] 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.

[36] Z. Mhowwala, A. Razia, and S. D. Shetty, "Movie rating prediction using ensemble learning algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 8, 2020, doi: 10.14569/ijacsa.2020.0110849.

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 PhD 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 reached via email at: lahcen.idouglid@uit.ac.ma.