STUDY OF FACTORS AFFECTING PROGRAM RATINGS IN WORKPOINT ENTERTAINMENT STATIONS VIA A MACHINE LEARNING ALGORITHM

WANPRA CHANTAWONGSATORN

Kasetsart University Laboratory School Multilingual Program Center for Educational Research and Development
E-mail: Wanpra2548@gmail.com

Abstract - Television influences the mindset of Thais. Thailand has comparable television viewing rates. The success of a television program depends on its ratings. Television stations' longevity and financial viability depended on advertising rates, which were affected by viewer competition. Therefore, channel content is essential. This study gathered data from Workpoint Entertainment Public Company Limited and inserted them into Machine Learning (ML) algorithms utilizing the RFR method. A total of 91 datasets. A schematic representation of the RFR algorithm was the ML algorithm used in this study. The accuracy of the RFR model was about 0.9310 and 0.9569 of R^2 for prediction and regression, respectively. From the prediction results, Programs were the most critical variable affecting the program's rating, 48%, followed by time and period, 15% and 10%, respectively. According to the Partial dependence report, the broadcast time of Workpoint Entertainment television stations with higher ratings should be broadcast in the range of 90 to 120 minutes, it can raise the rating, and from predictions, Two-dimensional partial, the broadcast date and the highest ratings of the program were Friday and Sunday at 7:00 p.m. Time slots from 10:00 a.m. to 4:00 p.m. do not affect the rating increase. In Random Forest Regression (RFR) in ML for the model, prediction helps forecast future new programs being produced on Workpoint Entertainment television stations. It can also be used for improvement and planning within the television station to be more successful in the programs.

Keywords - Artificial intelligence, Random Forest Regression (RFR), Rating, Television station management

I. INTRODUCTION

Currently, media is a necessity even though media can convey other people's feelings and thoughts. The primary method of communicating or transferring information, knowledge, and ideas among the public must be the use of the media. Television is a different sort of media that is accessible to learners and is located close to them. As a result, it is crucial in the current day [3]. The television industry is one of the highly competitive industries in today's competitive business situation. And they have intensified according to the development of technology that has changed to digital television. Including the rapid emergence of online media, changing the behaviour of viewers. Opting for two forms of daily communication and business operations [11]. Television media is a channel that greatly influences Thai people's thinking. It can be seen from the TV program ratings as high as 95% of the total population [1]. It is worth noting that the proportion of TV viewing in different regions of Thailand is similar. When TV media can reach a wide range of people. As a result, marketers see the importance of marketing communications for different target groups. Measuring popularity through TV ratings is an essential indicator of TV program success. To be a factor in selecting the most popular and influential TV programs and influence the selection of advertisements through this list [1]. A television program's material is rated to be appropriate for a specific age group of viewers. It's nothing new. This is because more than 30 nations employ television ratings, including Europe's continents Ireland and the United Kingdom. North American continents is Canada, and the United States. Oceania's continent is Australia. Africa South continent is Africa. Asia's continents are China, Vietnam, Malaysia, etc. [13]. It is a gauge of how well-liked TV series are based on how many viewers each station has. The viewers had stiff competition which influenced the popularity of watching various programs, which in turn influenced the advertising rate, a crucial element in ensuring TV stations' survival and financial success [14]. Therefore, the content of each channel is of the utmost importance. Attract viewers to come and watch the show [7].

Nonetheless, predicting how much material would boost a broadcast program's rating is impossible. In addition to the content, the program's ratings are also influenced by Other significant variables, such as the day and time a show airs, which can also impact. Since it is challenging to anticipate whether the ratings will rise or fall in advance, these predictions can be generated using machine learning (ML).

Technology development has advanced quickly in the current digital age. Technology is a significant force behind corporate growth and improves the overall quality of life [9]. ML is employed in industrial production processes to achieve satisfactory outcomes. ML may assist in issue-solving, and many organizations are using it to produce top-notch effects [4], and it also saves more time. ML instructs machines on how to manage data better. Sometimes,

even after reviewing the data, it's unable to understand the information that has been drawn from it. ML is widely used by a variety of businesses to their benefit [2]. ML is often utilized in a variety of jobs. Most of them are categorized objectively; email filtering, website rankings, diagnosis, face recognition, and many more examples are examples [6].

One such non-linear and non-parametric method is Random Forest Regression (RFR). It is a well-liked method for classifying things, regressing, choosing variables, identifying outliers, and determining the significance of variables. RFR is the term used to describe the usage of random forest for function approximation or regression. In RFR, numerous samples are randomly selected from a given data set, and Classification and Regression Trees (CART) are constructed. The outputs of each of these trees are eventually pooled, and an estimate of the target variables is created by averaging the outputs of each tree [5].

This research aims to use the RFR technique, which is part of ML, to predict the ratings of all 20 programs under Workpoint Entertainment data to make it helpful in predicting future new programs being produced. In addition, up in the Workpoint Entertainment television station can also be used to improve and plan within the television station.

The outline of this research is as follows. In section 1, information on television media in Thailand is presented. The importance of ratings affecting TV stations explains the importance and usefulness of ML and RFR techniques and their purpose. Section 2 shows the details of the input, output dataset, and RFR technique, including the model diagram and overview of this research. Section 3 presents the prediction accuracy results using all RFR techniques considered in this study. To find factors that affect program ratings of Workpoint Entertainment television stations. Finally, explain a brief conclusion in Section 4, which summarizes our findings.

II. MATERIALS AND METHODS

A. Data Collection

Data from Workpoint Entertainment Public Company Limited Pathum Thani, Thailand, were gathered for this study, and they were then substituted into an RF algorithm and viewed as being integrated into the Python environment, utilizing Scikit-learn libraries [12]. A total of 91 datasets, it consists of input data 13 variables and data output 1 variable composed of 20 programs represented by numbers 1 to 20, the months that the program aired in the past 3 months are September, October and November in 2022, represented by 9, 10 and 11, Studio for shooting represented by studio room numbers, studio room size (Area), broadcast date (Date) from Monday, Tuesday, Wednesday, Thursday, Friday, Saturday and Sunday, represented by numbers 1 to 7, broadcast time (Time), broadcast period (Period), the number of teams for shooting the program per 1 time, Budget for shooting a program per 1 time (Budget) (This is the cost of the master of ceremonies, stars in the program, committee, contestants, stylist, make-up, lights, LED cameras, sound, props, live band, cheerleaders, food, dancers/shows and copyright for images and music), Time to install and dismantle the scene (Scene) (in hours units), Time to install and dismantle Lighting (Lighting) (in hours units), Time to install and dismantle LED (LED) (in hours units), Duration in installing and dismantling Sound (Sound) (in hours units) were input data, and the average program rating in each month (Rating) was an output data as shown in Table 1. and an overview of the work is shown in Fig 1.

Number	Variables	Unit	Range	
			Minimum	Maximum
1	Programs	-	1	20
2	Months	-	9	11
3	Studio	-	1	19
4	Area	m^2	330	1677
5	Date	-	1	7
6	Time	a.m.	10.00	-
		p.m.	-	11.00
7	Period	min	60	120
8	Teams	-	67	125
9	Budget	Baht	400000	800000
10	Scene	hr.	12	20
11	Lighting	hr.	12	20
12	LED	hr.	6	18
13	Sound	hr.	6	10
14	Rating	-	0.156	3.9525

Table 1. Show the range of data collection from Workpoint Entertainment Public Company Limited.

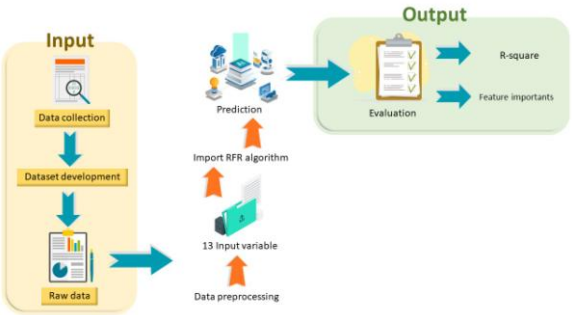


Fig 1. Research methodology

B. Random forest regression (RFR) algorithm

A schematic representation of the RFR algorithm, which was the ML algorithm used in this study, is shown in Fig 2. An ensemble model is the RFR. A model that increases prediction probability by mixing other models is called an ensemble model. In other words, compared to other single models, the ensemble model has greater predictive power. As seen in Fig. 2, the RFR prediction approach employs a regression tree to create an ideal model [13]. Bagging is a technique for leveraging a regression tree to drive the RFR algorithm. Bagging generates a tree from the data, after which the data are removed and restored. The technique mentioned above is then used to produce trees 2, 3, and N. Finally, the average of the trees is used to forecast anything. The study's usage of the RFR technique is common in numerical prediction analysis. Because the final result of the RFR method is derived by averaging out the most common values from the output of numerous decision trees, it will have a high level of accuracy.

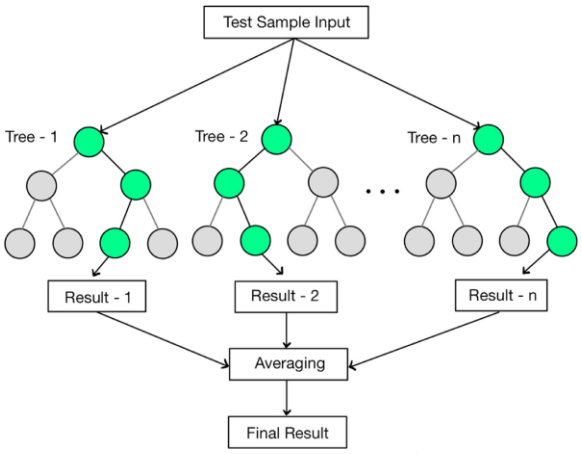


Fig 2. Diagrammatic representation of RFR [13].

III. RESULTS AND DISCUSSION

A. Pearson's analysis

A heatmap is a graph where the data values are represented by colours based on Pearson correlation. In other words, it uses colours to tell the reader something. When there is any data, using a heatmap is an excellent approach to direct the reader to the key regions. Heatmaps have an appealing visual appearance and convey simple data signals relatively instantaneously. Therefore, this method of correlation matrix visualization is used by both data scientists and data analysts. The heatmap produced is shown in Fig 3.

From the study of the coefficients that affect the interaction of programs by using ML to predict the results as shown in Fig 3. From the study, the coefficient can be divided into three parts: 1, 0, and -1, or can be seen from the colours of the coefficient: dark blue, white and dark red. The figure shows that the higher the correlation value, the darker the blue colour, indicating that the correlation is a direct variation. If the correlation is white, then the correlation is no effect. On the other hand, if the correlation is dark red, then the correlation is inverse variation.

From the correlation between LED and budget, there is a coefficient of 1, which is the highest coefficient. The result is dark blue. The result is dark blue, indicating LED takes a lot of time and will impact the cost of shooting one scene. This shows that these two variables are close. They were followed by the correlation between the lighting sound and scene sound with coefficients of 0.99 and 0.98, respectively. It shows that lighting can result in much time for sound. As a result, the sound is also high. In this section, it can have an indirect effect. That is, the budget used each time may also cause the budget to be increased. The correlations between studio rating, date-area, and lighting-studio coefficients of 0.23, 0.20, and 0.19 result in white. That means there was little or no interaction between these three. In other words, the studio does not affect the rating of the program, the date does not affect the studio, and lighting does not affect the studio or no interaction. Overall, teams, budget, time, scene, lighting, LED, and sound negatively affect the rating, thus making the resulting red. The coefficients were -0.29, -0.34, -0.36, -0.34, -0.34, and -0.3, respectively.

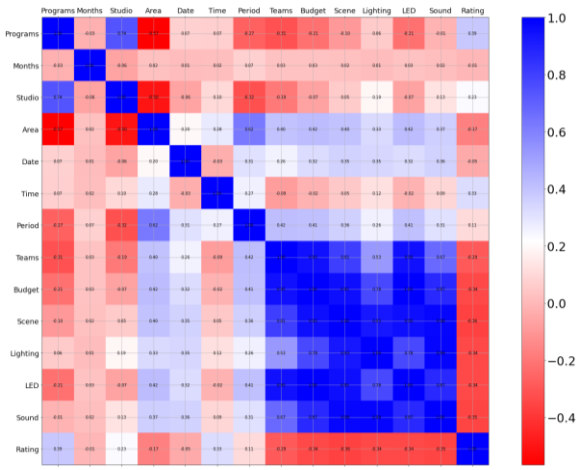


Fig 3. The correlation matrix coefficients affect the interaction for the rating of programs by using RFR to predict.

B. Feature evaluation

From the study of the effect of the interaction between performance and item that affects the television station rating. ML prediction results found that programs had the most significant impact on the show's rating at 48%. Followed by the time and period were 15% and 10%, respectively. However, other parameters can also affect a programs show's rating. These affect ratings less than the program, time, period, that is, LED, lighting, scene, budget, teams, area, and the studio also affect the item rating

is 3% as well. Followed by sound and date affecting the program's rating is 2%, and the month had the most negligible impact on program ratings by 1%, respectively. The results are shown in Fig 4.

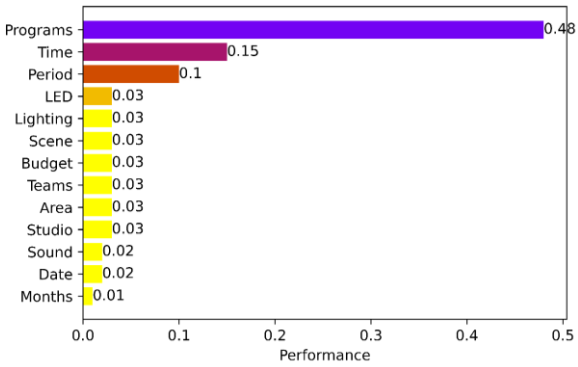


Fig 4. The effect of interaction between Performance and Item affects the television station rating.

C. Testing the model prediction accuracy

To better understand the prediction performance of the suggested model, Figure 5 compares instances of the dispersion of predicted values with actual (test) values for the ratings of Workpoint Entertainment television station during training and testing [12]. It was noted that the test dataset was from 10 folds of cross-validation for data sets of feature inputs. The straight line shows that the predictions and test data agree precisely [10]. The training case's R^2 values were within 0.9569, whereas the test case's R^2 for the Workpoint Entertainment television station's rating was 0.9310.

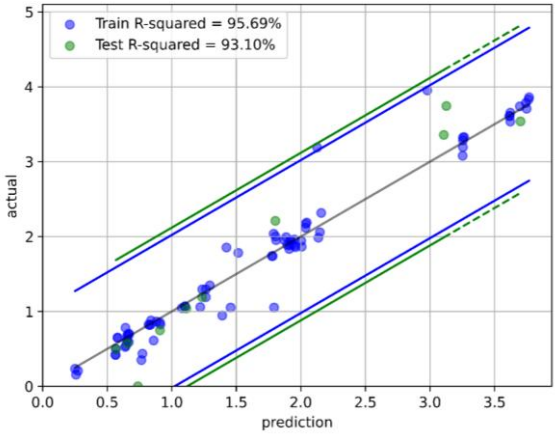


Fig 5. RFR ten-fold cross-validation test dataset prediction data distribution

D. Partial dependence

Fig 6. shows the interaction between the parameters. Period, time, and budget with rating. The graph indicates that the rating is not significantly impacted by the use of budget. According to the report, the time, which will boost the rating, should be between 6:00 p.m. and 8:00 p.m. If broadcasting after 8:00 p.m., the rating will decline, which is between 6:00 p.m. and 8:00 p.m., when all ages finish school or work, back home and it's a relaxing time for all genders and ages, thus making it the highest rating. According to the report, the duration of the period makes the program's rating even higher, which plays a significant role, if the television station would like to increase its rating, the duration of the broadcast should be 90-120 minutes, and it could raise the rating.

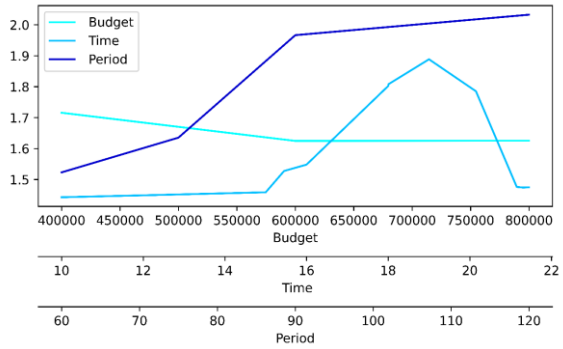


Fig 6. One-dimension partial dependence of the interaction between the parameters.

Fig 7. shows the results of the interaction between the date and time that affects the program's rating. The prediction shows that when broadcasting on Fridays and Sundays, the average rating will increase to the highest, 1.89 at 7:00 PM, due to the time off work and weekends. It was also a time for everyone in the family to relax. It is, therefore a part that can increase the rating. From 4:00 p.m. to 6:00 p.m. from Monday to Sunday, the predicted ratings were 1.52, 1.59, 1.67, and 1.74, respectively. This increase in rating looks the same from Monday to Sunday. The rating did not increase from 10:00 a.m. to 4:00 p.m. because it was when everyone was working or studying. Therefore, if a television station wants to increase its rating, it should broadcast on Fridays and Sundays at 7:00 p.m.

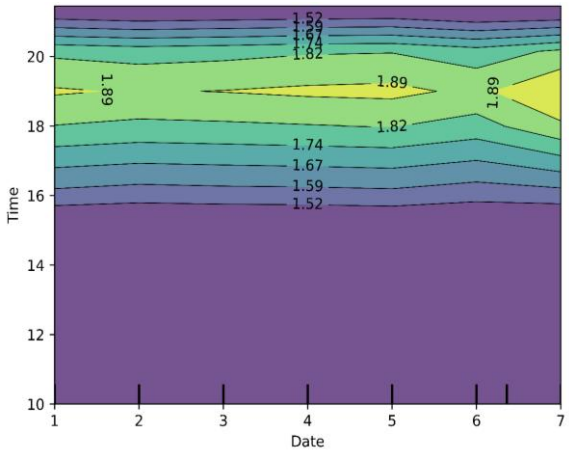


Fig 7. Two-dimension partial dependence of the interaction between the date and the time that affects the program's rating.

IV. CONCLUSION

Workpoint Entertainment is a television station in Thailand. The general program's format emphasis is mainly on game shows and variety shows. The RFR algorithm was applied to model and predict the Workpoint Entertainment television program during September, October, and November 2022. It was using input features from 13 variables as input data. This includes group 1: people datasets (teams, budget, studio, and area); group 2: demolition installation datasets (scene, lighting, LED, and sound); group 3: broadcast datasets (programs, months, date, time, and period); and output one variable is rating.

The accuracy of the RFR model was about 93.10% and 95.69% of R^2 for prediction and regression, respectively. From the prediction results, programs were the most important variable that affects the program's rating, at 0.48, followed by time and period, at 0.15 and 0.1, respectively. From a study of partial dependence, the broadcast duration of 90 to 120 minutes will increase the ratings of Workpoint Entertainment programs maximum increase. From predictions, from predictions, Two-dimensional partial, at 7:00 p.m. on Friday and Sunday. It is the period that makes the ratings of the program increase to the maximum. Prediction is helpful in RFR for the model in predicting upcoming new programs produced by Workpoint Entertainment television stations. It may also be utilized for planning and enhancement within the television station to enhance the success of the programming. Prediction is helpful in RFR for the model in predicting upcoming new programs produced by Workpoint Entertainment television stations. It may also be utilized for planning and enhancement within the television station to enhance the success of the programming.

ACKNOWLEDGEMENT

I would like to thank the Kasetsart University Laboratory School Multilingual Program Center for Educational Research and Development for supporting research. I would like to express my gratitude for the help with my research paper. This work was supported by Workpoint Entertainment Public Company Limited in Pathum Thani, Thailand, which provided information for conducting this research.

Lastly, I appreciate my parents, siblings, and support system, who never cease to love and support me. They believed in me, and as a result, I achieved success. I will always be grateful for the unwavering love and support I received every day and during the entire research process. Without them, my study could not have been possible.

REFERENCES

[1] B. Chatsawvapan, “Social Media Marketing Communication Strategies For Television Programs,” Master of Arts (Communication Arts and Innovation)., pp. 1-132

[2] B. Mahesh, “Machine Learning Algorithms - A Review,” International Journal of Science and Research (IJSR)., pp.381-386, Jan 2018

[3] P. Namboonruang, "Television," Technology, innovation, invention, engineering, agriculture.

[4] R. Maythachalermphat, "Applying Machine Learning to Industrial Work (Part 1)," Engineer, Cyber-Physical Systems (CPS) Team IoT and Industrial Automation Research Group (IIARG) Nectech NSTDA., Aug 2019.

[5] S. Desai and Taha B.M.J. Ouarda, "Regional hydrological frequency analysis at ungauged sites with random forest regression," Journal of Hydrology., vol. 594, no. 125861, March 2021.

[6] S. Jamal, S. Goyal, A. Grover, and A. Shanker, "Machine Learning: What, Why, and How?," Bioinformatics: Sequences, Structures, Phylogeny., pp. 359–374, Oct 2018.

[7] S. Khamthong, "An Analysis of the Survival of the Digital TV Industry," Office of the Broadcasting Commission television business and the National Telecommunications Commission., pp.136-157, 2017.

[8] S. Kwak et al., “Machine learning prediction of the mechanical properties of g-TiAl alloys produced using random forest regression model,” Journal of materials research and technology., vol 18, pp. 520-530, May–June 2022.

[9] S. Nonsiri, "What is the Python programming language," 9Expert Training.

[10] S. Phromphithak, T. Onsree, and N. Tippayawong, “Machine learning prediction of cellulose-rich materials from biomass pretreatment with ionic liquid solvents,” Bioresource Technology., vol 323, no. 124642, March 2021.

[11] S. Yitansi, "Co-Production TV Production Innovation in Thai Television IndustryThe Case Study of Workpoint Entertainment Public Company Limited," Siam Communication Review., vol.2, pp. 28-44, Nov 2021.

[12] T. Katongtung, T. Onsree, and N. Tippayawong, "Machine learning prediction of biocrude yields and higher heating values from hydrothermal liquefaction of wet biomass and wastes," Bioresource Technology., vol. 344, Part B, no. 126278, Jan 2022.

[13] Thai Health Report 2008, "Institute for Population and Social Research, Mahidol University Thai Health Promotion Foundation (Thai Health Promotion Foundation)., pp. 71-75, 2008.

[14] W. In-iad, "Guidelines for supervising the presentation of television drama content that is inappropriate to children and youth," Office of the National Broadcasting Commission television business and the National Telecommunications Commission., pp. 290-315, 2017.

★★★