

Television Advertising Slot Purchase Decision-Making Framework

Zaem ZAINAL and Seiya HASHIZUME

Lifenet Insurance Company 8th floor, Nibancho Center Building, 5-25 Niban-cho, Chiyoda-ku, Tokyo 102-0084,
Japan

E-mail: {zaem.zainal, seiya.hashizume}@lifenet-seimei.co.jp

Abstract Television (TV) advertising generates immediate, measurable spikes in online consumer activity. However, selecting the most effective advertising slots prior to purchase remains a significant challenge for advertisers. This study proposes a decision-making framework designed to evaluate and compare potential TV advertising slots using empirical historical data. To integrate disparate datasets - specifically, agency quotation lists and internal web traffic logs- we applied the Jaro-Winkler similarity metric. This approach achieved a match rate of 89.9%, a substantial improvement over the 47.4% rate obtained through exact string matching. A LightGBM model was developed to predict web traffic spikes. The model achieved a 42.6% Mean Absolute Percentage Error (MAPE), significantly outperforming linear regression models. This framework enables advertisers to objectively compare potential slot investments based on predicted performance indicators and audience metrics.

Keyword Television Advertising, Web Traffic, Decision-Making Framework, E-commerce Performance, Jaro-Winkler Similarity, Generative AI

1. Introduction

There are many ways to influence consumer behavior through advertising; however, television (TV) advertising is particularly valuable for e-commerce business models due to its ability to generate immediate and measurable data. Previous studies reported a noticeable spike in online activity immediately following a broadcast [1,2]. Figure 1 illustrates the spike in web traffic to an e-commerce website following a TV commercial broadcast at 20:50.

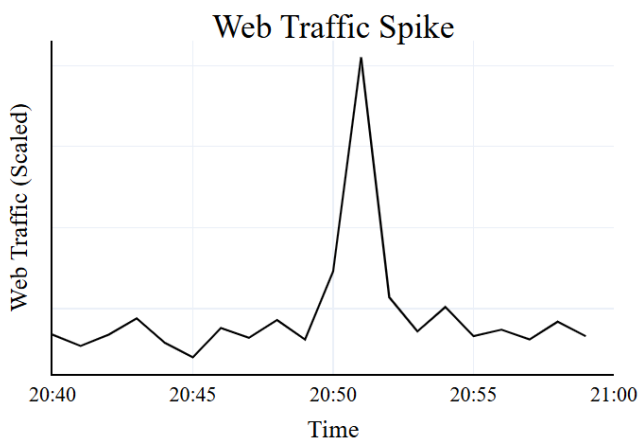


Figure 1. Example of a web traffic spike (scaled)

While existing studies evaluate the impact of TV advertising on online activity (or engagement) [1,2,3], their findings are not directly applicable to the practical decision-making process involved in purchasing specific advertising slots.

In general, the purchase of a TV advertising slot proceeds

as follows. The advertiser receives a quotation list of several slots from advertising agencies based on desired conditions such as price range, proposed Gross Rating Points (GRP), and preferred broadcasting time. Determining the superior return on investment is simple when comparing two slots, but the task becomes increasingly difficult when evaluating a multitude of competing options. We propose a framework that scales beyond simple comparisons to support complex, data-driven purchasing decisions.

Our approach addresses two primary challenges inherent in this real-world problem. First, we construct a specialized dataset by integrating slot information from advertising agencies with empirical web traffic data provided by Lifenet Insurance Company. Second, we develop a decision-support framework that utilizes machine learning to predict performance indicators. By providing actionable metrics based on these predictive models, the proposed framework assists decision-makers in evaluating the viability of each advertising slot, ultimately facilitating more efficient and effective investment strategies.

2. Framework for Data Integration

As outlined in the introduction, the primary challenge is to integrate two separate data sources: (a) monthly order lists submitted to advertisement agencies and (b) post-advertisement web traffic spike data (henceforth referred to as spike data) recorded internally from January 2023 to June 2025. In addition to spelling inconsistencies and notes due to manual data entry, when compared to slots in (a)

order list, some slots in (b) spike data had a change in TV program names, a change in broadcast time, or a combination of both. To illustrate the challenges in this paper, we present examples using imaginary program names. In (a) order list the program name might be “Action Animation”, but in (b) spike data it might be recorded as “Action Animation Alpha”. This makes it difficult to match the two datasets using TV program name strings as the key. Exact matching on {broadcast station, broadcast date, TV program name} only accounted for 48.4% of the data in (b). Similarly, using TV program broadcast time (e.g. 15:00~16:00) instead of name only accounts for 57.2% of the data. To overcome this, we propose cross matching using both TV program name and broadcast time, in addition to Jaro-Winkler similarity to find matching strings with the highest similarity.

2.1. Jaro-Winkler Similarity

The Jaro-Winkler similarity is a string metric designed to measure the edit distance between two sequences. It is an extension of the Jaro similarity metric, incorporating a prefix scale that gives more favorable ratings to strings with common prefixes. The Jaro similarity between two strings s_1 and s_2 is defined as:

$$sim_{Jaro}(s_1, s_2) = \begin{cases} 0 & \text{if } m = 0 \\ \left(\frac{1}{3}\right) \left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m}\right) & \text{otherwise} \end{cases}$$

where:

- $|s_i|$ denotes the length of string s_i
- m is the number of matching characters
- t is half the number of matching characters that are out of order

Two characters from s_1 and s_2 are considered matching

if they are identical and not farther than $\left\lfloor \frac{\max(|s_1|, |s_2|)}{2} \right\rfloor - 1$

positions apart. Transpositions are defined as the number of matching characters that are in different sequence orders divided by two.

The Jaro-Winkler similarity incorporates a prefix scale p that gives additional weight to strings that match from the beginning:

$$sim_{JW}(s_1, s_2) = sim_{Jaro}(s_1, s_2) + l \cdot p \left(1 - sim_{Jaro}(s_1, s_2)\right)$$

where:

- l is the length of the common prefix (set as two characters in this study)
 - p is a scaling factor for how much the score is adjusted upward for having common prefixes.

The Jaro-Winkler similarity produces values in the range [0,1], where 1 indicates an exact match and 0 indicates no similarity. This metric has been widely adopted in record linkage and duplicate detection tasks due to its effectiveness in handling typographical errors and variations in name spellings. To increase the match rate, we reversed the string to apply Jaro-Winkler similarity on the suffix of the string as well, which has been adopted in a previous study [4]. The final score of a match is the higher of the two values. We used Jaro-Winkler similarity when matching on different keys resulted in a conflict between two records. The final flow of processing for matching data from (a) order list on (b) spike data:

2.2. Matching Procedure

We integrated the datasets using the following 5 steps:

- **Step 1:** Perform exact matching using a composite key of {broadcast station, date, program name}.
- **Step 2:** For unmatched records, perform matching based on {broadcast station, date, broadcast time}.
- **Step 3:** If multiple candidates emerge in Step 2, the record with the highest Jaro-Winkler similarity for the program name is selected.
- **Step 4:** For remaining unmatched records, match where the commercial's specific broadcast time in (a) falls within the start and end times of the program in (b).
- **Step 5:** If multiple candidates emerge in Step 4, the record with the highest Jaro-Winkler similarity is selected.

This resulted in a match rate of 47.4% at Step 1, 60.9% at Step 3, and a final match rate of 89.9% of the data at Step 5. On closer inspection, those which did not match were time slots whose programming was changed entirely.

3. Model Implementation

Using dataset created in the previous section, we proceed to create a prediction model for our decision-making framework. Since most of the TV advertisements guide the user to Lifenet Insurance Company’s insurance plan quotation page, this study will focus on the spike of unique users visiting the page as the key indicator of a TV advertisement slot’s performance. Other online activity was not included in this study. For simplicity we refer to this as “quotation spike”. Input data includes broadcasting time, proposed GRP, M1 proposed rating (viewer rating for males aged 20-34), F1 proposed rating (viewer rating for

females aged 20-34), broadcasting station and advertisement creative content.

3.1 Feature Creation with Generative Artificial Intelligence (AI)

We attempted to use generative AI to create additional features from existing data for machine learning. In a previous study, facial tracking software was used to examine how emotional context of a TV program impacts TV advertisement effectiveness [5]. Six extra features were created to reflect the emotional context of a TV program: healing, excitement, cognitive load, tempo, background viewing and co-viewing. Each was evaluated on a scale of 1 to 10 by using generative AI with TV program names as input, inferred from program genres and metadata. Examples of imaginary program name inputs and their respective emotional context output are shown in Table 1.

Table 1. Emotional context is inferred from imaginary TV program names by the generative AI.

Program Name	Healing	Excitement	Cognitive Load	Tempo	Background Viewing	Co-viewing
Action Animation	3	9	5	9	4	8
Cooking Hour	7	5	4	5	8	7
News Live	3	6	7	7	8	6

“Action Animation” has a high excitement and tempo score, whereas “News Live” has a higher cognitive load score.

3.2 Machine Learning

For machine learning, we used Light Gradient-Boosting Machine (LightGBM), an ensemble learning framework that uses gradient boosting method [6]. We compared results with a simple linear regression and multiple linear regression model. The last two months of the dataset (July and August 2025) were used as test data. Models were evaluated using Mean Absolute Percentage Error (MAPE)

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right|$$

where:

- n is the number of observations (data points).
- \sum adds up the values for all i .
- A_i is the actual value.

- F_i is the predicted value.

On evaluation, LightGBM yielded a MAPE of 42.6% on test data, whereas simple and multiple linear regression models yielded 58.0% and 53.8% respectively. Top important features include proposed GRP, advertisement creative content (encoded to average spike per content type) and TV advertisement slot start time. Emotional context created using generative AI did not have a relatively high feature importance compared to the most important feature: proposed GRP.

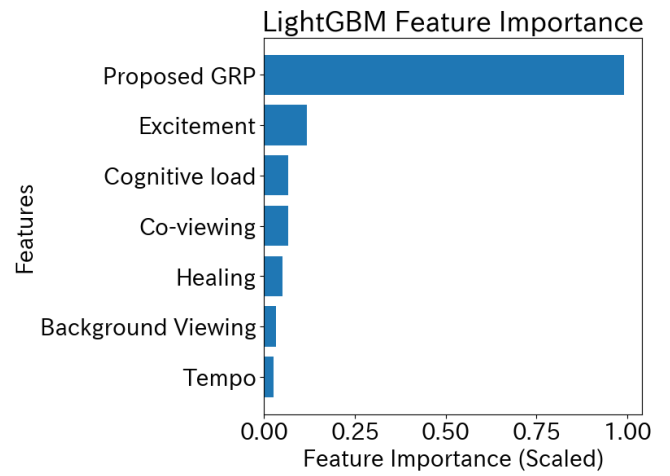


Figure 2: Feature importance of emotional context in comparison to proposed GRP

4. Discussion

In the process of creating the dataset, approximately 10% of data remained unmatched due to a change in programming, as purchasing often occurs a month in advance. In addition to TV program names for matching, which fails in this instance, matching the remaining 10% based on the date and time of the TV program slot is required.

The proposed decision-making framework based on LightGBM still has a high MAPE value which can be improved on. Nearly three years of data was used for the model. However, changes in trends and advertisement creative content presents a challenge when selecting a suitable data range, since there is a trade-off between having more relevant recent data and sample size. In order to provide a viable solution for this “real-world” problem, our future works will focus on deriving stable conclusions from limited data sets. The presence of distinct time-series trends is the direct cause of the limited usable dataset size. Therefore, retraining the model with the most recent data

possible will have a significant impact on accuracy.

Proposed GRP and creative content were the dominant predictors for the model. This also made the model prone to overfitting. To mitigate this, training error was adjusted by tuning parameters such as tree depth. In addition, feature-engineering these two dominant variables is important to enhance model performance. Specifically, converting creative content into numerical representations remains a perpetual challenge. For example, quantifying creative content in advance through monitor feedback or surveys before analysis may improve prediction accuracy.

For cases with low GRP (low impact), noise from organic traffic unrelated to TV advertisements is significant. While optimizing with Root Mean Squared Error (RMSE) is desirable to address this, the small sample size makes the model vulnerable to outliers—a known downside of RMSE. If more data becomes available, building separate models for low-GRP and high-GRP segments may be effective.

Although using generative AI on TV show names to create features did not result in high-importance features, of the six extra features, excitement came out on top. This may suggest that TV programs with higher engagement have a larger impact on TV advertisement effectiveness.

5. Conclusion

This study provides a TV advertising slot purchase decision-making framework based on a dataset created from two different data sources combined using Jaro-Winkler similarity, which yielded an 89.9% match rate compared to an exact matching method. LightGBM was used to predict spikes of TV advertisement slots with a MAPE of 42.6%. For actual implementation, there is still improvement required in accuracy. Specifically, we intend to reduce the effect of noise from traffic unrelated to TV advertisements, as outlined in the discussion section.

References

- [1] L. Veverka and V. Holý, "The Impact of TV Advertising on Website Traffic," *Appl. Stochastic Models Bus. Ind.*, 2024, Art. no. 1, Advance online publication.
- [2] R. Y. Du, M. Joo, and K. C. Wilbur, "Immediate Responses of Online Brand Search and Price Search to TV Ads," *J. Market.*, vol. 83, no. 4, pp. 81–100, 2019.
- [3] J. Liukonyte, T. Teixeira, and K. C. Wilbur, "Television Advertising and Online Shopping," *Market. Sci.*, vol. 34, no. 3, pp. 311–330, 2015.
- [4] K. B. K. Malaga, K. L. Verdillo, and E. S. Pascual, "An Enhancement of the Jaro-Winkler Fuzzy Searching Algorithm Applied in Library Search Engine," *J. Inf. Syst. Eng. Manage.*, vol. 10, no. 28s,

pp. 649–660, 2025.

- [5] .S. Bellman, B. Wooley, and D. Varan, "Program-Ad Matching and Television Ad Effectiveness: A Re inquiry Using Facial Tracking Software," *J. Advert.*, vol. 45, no. 1, pp. 72–77, 2016.
- [6] G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 3146–3154.