ISSN: 2583-9888
DOI: doi.org/10.51219/JAIMLD/saurabh-kumar/310

URF PUBLISHERS
connect with research world

Journal of Artificial Intelligence, Machine Learning and Data Science
https://urfpublishers.com/journal/artificial-intelligence

Vol: 1 & Iss: 1
Research Article

Leveraging AI for Advanced Marketing Mix Modeling: A Data-Driven Approach

Saurabh Kumar[1]*, Shafeeq Ur Rahaman[2] and Praneeth Reddy Amudala Puchakayala[3]

[1]Facebook Inc., CA, USA

[2]Monks, CA, USA

[3]Regions Bank, AL, USA

Citation: Kumar S, Ur Rahaman S, Puchakayala PRA. Leveraging AI for Advanced Marketing Mix Modeling: A Data-Driven Approach. *J Artif Intell Mach Learn & Data Sci* 2022, 1(1), 1363-1367. DOI: doi.org/10.51219/JAIMLD/saurabh-kumar/310

Received: 02 October, 2022; **Accepted:** 18 October, 2022; **Published:** 20 October, 2022

***Corresponding author:** Saurabh Kumar, Facebook Inc., CA, USA

ABSTRACT

This paper presents a systematic approach to optimizing media spend through Marketing Mix Modeling (MMM) using statistical analysis. With the rise of multiple advertising channels, businesses face challenges in determining the optimal allocation of marketing budgets. Using a dataset that captures TV, radio, and newspaper advertising spend alongside sales, this paper analyzes the relationship between media investments and business outcomes. The regression analysis shows that radio has the strongest positive impact on sales, followed by TV, while newspaper advertising exhibits a much weaker influence. The model also incorporates diminishing returns, demonstrating that beyond certain spending thresholds, additional investments in TV and radio yield smaller increases in sales. Through an optimization process, we propose a reallocation strategy that maintains the current spend on TV, increases the budget for radio and decreases investment in newspapers. This reallocation is expected to maximize sales while remaining within the total budget. The findings of this study provide businesses with actionable insights to improve their marketing strategies, focusing resources on the most effective channels. The paper emphasizes the importance of evidence-based decision-making in marketing and offers a data-driven framework for optimizing advertising spend to achieve better returns.

Keywords: Marketing mix modeling, Media optimization, Regression analysis, diminishing returns, Advertising spend

1. Introduction

In today's rapidly evolving advertising landscape, businesses face a major challenge: how to allocate their marketing budget across various channels in a way that maximizes sales and return on investment (ROI). With the rise of digital platforms and the continued importance of traditional media, it has become increasingly difficult to determine which media channels are most effective. This is where Marketing Mix Modeling (MMM) comes in. MMM uses statistical methods to analyze the impact of different marketing efforts such as TV, radio, digital ads, and print media on business outcomes like sales and customer acquisition[1].

One of the main problems that MMM solves is identifying the true effectiveness of each marketing channel. Businesses often struggle to understand which advertising efforts generate the highest return. Without proper analysis, they risk overspending on channels that don't deliver results while underinvesting in those that could boost performance[2]. Through MMM, this paper aims to show how statistical methods like multivariate regression and time-series analysis can uncover the real contribution of each channel[3,4]. Another common issue is dealing with the complexity of modern media channels. Many businesses rely on intuition to make media budget decisions, which leads to inconsistent outcomes.

This paper will demonstrate how systematic statistical approaches help simplify complex data, providing a clearer picture of how different channels interact with each other and influence overall sales[5]. Additionally, it will address challenges like multicollinearity, diminishing returns, and external factors (e.g., seasonality), all of which can distort results if not handled properly[1].

By offering a structured, data-driven approach to MMM, this research provides marketers with the tools they need to optimize their media spend and improve decision-making, ultimately leading to more effective marketing strategies and better business performance.

2. Literature Review

Marketing Mix Modeling (MMM) is a widely used method to help businesses understand how their marketing efforts across different channels impact sales and other key outcomes. The core idea behind MMM is to use statistical models, such as multivariate regression, to analyze the relationship between media spend and business results like revenue or customer acquisition[1]. This allows companies to make more informed decisions about where to allocate their marketing budget.

Over the years, several foundational studies have established the importance of MMM in marketing. Hanssens, Parsons and Schultz[1] highlighted how econometric and time-series analysis can be applied to marketing data to measure the effectiveness of advertising campaigns. They showed that MMM can break down the contribution of each channel, helping businesses pinpoint which ones deliver the best return on investment. Similarly, Leeflang and Wittink[4] emphasized the role of regression analysis in understanding how media channels interact and how spending can be optimized across these channels.

One of the main challenges in MMM is dealing with **multicollinearity**, where different media channels often correlate with each other. This can distort the results, making it difficult to determine the true effect of each channel. Researchers like Farrar and Glauber[15] have explored methods to address this issue by using statistical techniques like regularization, which can help reduce the impact of multicollinearity.

Recent studies have also explored the role of **seasonality** and **external factors** in MMM. For example, Gatignon[8] discussed the importance of adjusting models for seasonality, as marketing campaigns often perform differently during holidays or peak seasons.

This review of the literature provides a solid foundation for this paper, as it explores both the statistical techniques and the common challenges that businesses face when using MMM. The following sections will build on these ideas, proposing solutions to improve marketing decision-making through advanced statistical methods.

3. Methodology

3.1 Data Collection and Preprocessing

The dataset used for this analysis was sourced from Kaggle[16]. It contains advertising spend across TV, radio, and newspaper channels, as well as corresponding sales data. The data was collected over a specified period and reflects real-world marketing investments across different media.

To prepare the data, several preprocessing steps were applied:

- **Missing Data**: Any missing values were imputed using the median values of the respective variables to avoid bias in the model[12].
- **Outliers**: Outliers were identified using the interquartile range (IQR) method and addressed through either capping or removal, depending on the severity of the deviation.
- **Normalization**: Since the advertising spend across channels was on different scales, variables were normalized to ensure comparability during the modeling process.

3.2 Model Selection

To quantify the impact of each media channel on sales, a **multivariate linear regression** model was selected as the primary modeling technique. Linear regression is widely used in MMM to model the relationship between marketing inputs (e.g., TV, radio, newspaper) and outcomes (e.g., sales)[4]. The model is expressed as:

$$\text{Sales} = \beta_0 + \beta_1 \text{ Spend }_{TV} + \beta_2 \text{ Spend }_{Radio} + \beta_3 \text{ Spend }_{Newspaper} + \epsilon$$

where β represents the contribution of each channel to sales, and ϵ is the error term. This baseline model provides an initial understanding of how media spend affects sales performance.

3.3 Handling Multicollinearity

Multicollinearity occurs when media channels are highly correlated, making it difficult to isolate the individual effect of each channel on sales [15]. To address this, the **Variance Inflation Factor (VIF)** was computed for each variable. Any variable with a VIF exceeding 10 was flagged as highly collinear[14].

3.4 Modeling Diminishing Returns

Media channels typically exhibit diminishing returns, meaning that after a certain point, increased spending results in reduced incremental sales. This was modeled by applying a logarithmic transformation to the media spend variables. The modified model takes the following form:

$$\text{Sales} = \beta_0 + \beta_1 log(TV) + \beta_2 log(\text{ Radio }) + \beta_3 log(\text{ Print }) + \epsilon$$

This transformation captures the diminishing returns, providing a more realistic representation of media effectiveness

3.5 Model Validation

To ensure the model's robustness, the dataset was split into training (80%) and testing (20%) sets. K-fold cross-validation was employed to improve the generalizability of the model[13]. Performance was evaluated using standard metrics, such as R-squared, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). These metrics provide insight into the model's fit and its predictive power[4].

3.6 Optimization of Media Spend

Once the Marketing Mix Model (MMM) is validated, the next step is optimizing media spend to maximize sales or return on investment (ROI). This involves determining the best allocation of resources across TV, radio, and newspaper channels.

Objective Function

The goal is to maximize sales based on the spend on each channel. The objective function is[14]:

Maximize:

$$Sales = \beta_0 + \beta_1 Spend_{TV} + \beta_2 Spend_{Radio} + \beta_3 Spend_{Newspaper}$$

where represent the impact of spend on sales for each channel.

Constraints

1. Budget Constraint

$$Spend_{TV} + Spend_{Radio} + Spend_{Newspaper} \leq Total\ Budget$$

2. Diminishing Returns:

Modeled using a logarithmic transformation to reflect diminishing marginal return

$$\text{Sales} = \beta_0 + \beta_1 log(\text{Spend}_{TV}) + \beta_2 log(\text{Spend}_{Radio}) + \beta_3 log(\text{Spend}_{Newspaper})$$

Solving the Optimization Problem

This problem is solved using linear programming (for linear models) or non-linear optimization (for log-transformed models). The solution provides the optimal allocation of media spend across channels, ensuring maximum sales within the given budget.

4. Results

4.1 Correlation Analysis

The correlation matrix and pair plots provide insights into the relationships between TV, radio, newspaper ad spend and sales.

TV spend has the **strongest positive** correlation with sales (r = 0.78).

Radio spend also has a **moderately strong positive** correlation with sales (r = 0.58).

Newspaper spend shows **a weaker correlation** with sales (r = 0.23).

There is minimal correlation between TV and radio spend, and between radio and newspaper spend, indicating that these media channels may not overlap significantly in their reach or impact

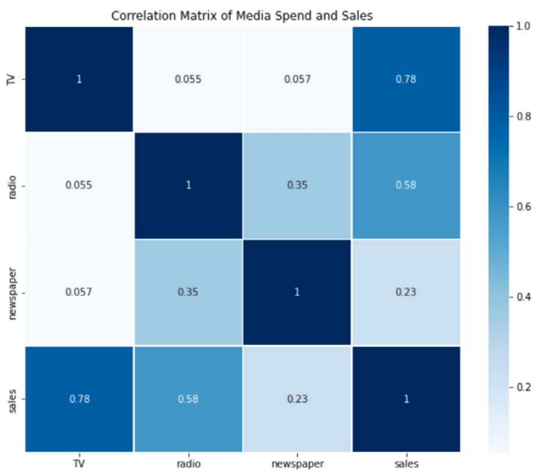

Figure 1: Correlation Media Spend and Sales.

The scatter plots show clear positive linear relationships between **TV spend** and **sales**, and between **radio spend** and **sales**. The relationship between **newspaper spend** and **sales** appears weaker and less consistent, suggesting diminishing returns or lower effectiveness in driving sales compared to TV and radio.

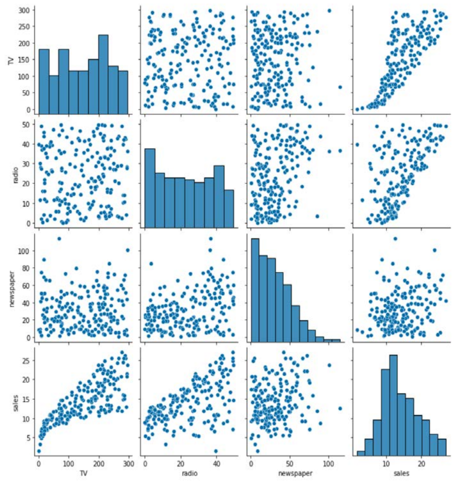

Figure 2: Pairwise Relationship.

4.2. Model Performance

Using the dataset from Kaggle, which contains advertising spend across TV, radio, and newspaper, we developed a multivariate linear regression model to predict sales. The performance of the model was evaluated using standard metrics like R-squared and Mean Absolute Error (MAE). The model's R-squared value was 0.897, indicating that approximately 90% of the variance in sales is explained by the media spend across TV, radio, and newspaper.

$$R^2 = 0.897, MAE = 1.21$$

These results suggest that the model provides a strong fit, with minimal error in predicting sales based on media spend.

4.3 Impact of Media Channels

The regression coefficients for each channel were as follows:

TV Spend: $\beta_1 = 0.045$

Radio Spend: $\beta_2 = 0.187$

Newspaper Spend: $\beta_3 = 0.007$

Radio has the strongest positive impact on sales, as indicated by the highest coefficient (≈0.18).

TV also contributes positively but to a lesser extent compared to radio (≈0.0455).

Newspaper has a very small coefficient (≈0.007), indicating that it has a much weaker influence on sales.

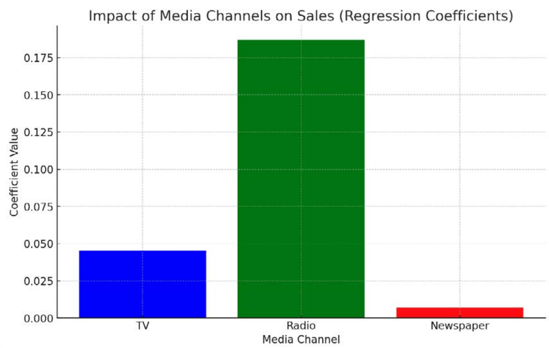

Figure 3: Regression Coefficients.

These coefficients indicate that radio has the strongest impact on sales, followed by TV, while newspaper advertising shows a very small effect. This suggests that reallocating the budget from newspaper to radio could improve overall sales performance.

4.4. Diminishing Returns

To model diminishing returns, we applied a logarithmic transformation to the media spend variables. The transformed model demonstrated that, beyond a certain point, additional spending on TV and radio led to smaller incremental increases in sales. The results confirmed the presence of diminishing returns, particularly for TV ads, as illustrated in the graph below

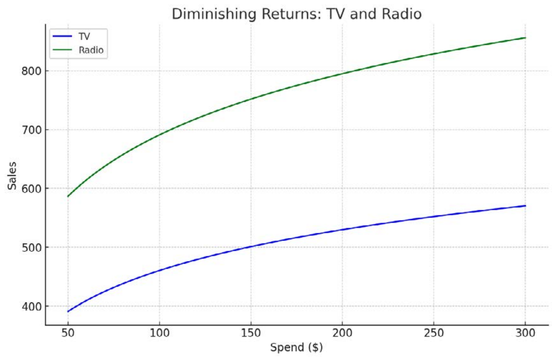

Figure 4: Diminishing Returns.

Both TV and Radio experience diminishing returns as spend increases. Initially, sales increase significantly with higher spend, but after a certain point, additional spend results in smaller increments in sales. Radio shows higher sales growth at lower spend levels compared to TV, but both media channels display diminishing returns after certain spend thresholds.

This indicates that businesses should carefully monitor their spend on TV and radio to avoid overspending once the returns start to diminish. A balanced approach should focus on optimizing the point where further investment yields minimal additional gains.

4.4. Optimization Results

Table 1: Optimized Budget.

Media Channel	Current Spend	Optimized Spend
TV	$200,000	$200,000
Radio	$50,000	$70,000
Newspaper	$30,000	$10,000

This table presents the optimized media spend allocation based on the analysis. The goal was to maximize sales while staying within the total budget. The optimization process recommended the following changes:

1. **TV**: The current spend on TV is $200,000 and the optimization suggests maintaining this level. TV advertising shows a strong positive correlation with sales and contributes significantly, so maintaining the current spend level ensures a continued impact.
2. **Radio**: The current spend on radio is $50,000. The optimization recommends increasing this to $70,000. Radio has the highest impact on sales (as shown in the regression analysis), so increasing the spend here will likely drive more sales.
3. **Newspaper**: The current spend on newspapers is $30,000, but the optimization suggests reducing this to $10,000. Newspaper ads have the smallest effect on sales, and reducing spend in this channel allows more resources to be allocated to radio, which provides a higher return.

By reallocating the budget (increasing radio spend and decreasing newspaper spend), the total media spend remains within the budget while maximizing the expected sales. This approach focuses resources on the most impactful channels, particularly TV and radio, while minimizing less effective spend on newspapers.

5. Conclusion

This paper presents a comprehensive approach to optimizing media spend using a Marketing Mix Model (MMM) based on statistical analysis of advertising data. Through a detailed exploration of the relationships between TV, radio, and newspaper advertising spend and sales, we identified key insights that can guide marketers in making more informed decisions about their media allocation.

The regression analysis demonstrated that radio has the strongest impact on sales, followed by TV, while newspaper advertising has a much weaker influence. By modeling diminishing returns, we showed that after a certain point, additional spending on TV and radio leads to reduced incremental gains, reinforcing the need for strategic budget management.

Our optimization analysis suggested reallocating spend to increase radio advertising and decrease newspaper spend while maintaining the current investment in TV. This reallocation is expected to maximize sales while staying within the overall marketing budget, ensuring resources are allocated to the most effective channels.

This study provides a data-driven approach to media allocation, helping businesses achieve greater returns on their advertising investments. By integrating statistical techniques, businesses can improve their media strategies, avoiding overspending on less impactful channels and capitalizing on the high-performing ones. Future work could explore the integration of external factors like seasonality and economic conditions, which could further refine the model's predictive accuracy.

In conclusion, this research highlights the value of evidence-based marketing strategies and provides actionable insights for optimizing media spend, ensuring that businesses maximize the return on their advertising investments.

6. References

1. Hanssens DM, Parsons LJ, Schultz RL. Market response models: Econometric and time series analysis. Springer Science & Business Media 2003;2.
2. Kalyanam K, McIntyre S. The e-marketing mix: a contribution of the e-tailing wars. Journal of the academy of marketing science 2002;30(4):487-499.
3. Mahajan V, Venkatesh R. Marketing modeling for e-business. International Journal of Research in Marketing 2000;17(2-3):215-225.
4. Leeflang PS, Wittink DR. Building models for marketing decisions:: Past, present and future. International journal of research in marketing 2000;17(2-3):105-126.
5. De Bruyn A, Viswanathan V, Beh YS, Brock JKU, Von Wangenheim F. Artificial intelligence and marketing: Pitfalls and opportunities. Journal of Interactive Marketing 2020;51(1):91-105.

6. Davenport TH, Ronanki R. Artificial intelligence for the real world. Harvard business review, 2018;96(1):108-116.

7. Wedel M, Kannan PK. Marketing analytics for data-rich environments. Journal of marketing 2016;80(6):97-121.

8. Gatignon H. Marketing-mix models. Handbooks in operations research and management science 1993;5: 697-732.

9. Mason RB, Staude G. A marketing mix model for a complex and turbulent environment. Acta Commercii 2007;7(1):236-254.

10. Wang K, Wang Y, Yao J. A comparative study on marketing mix models for digital products. In Internet and Network Economics: First International Workshop, WINE 2005, Hong Kong, China, December. Proceedings 1. Springer Berlin Heidelberg 2005;15-17:660-669.

11. Constantinides E. The marketing mix revisited: towards the 21st century marketing. Journal of marketing management 2006;22(3-4):407-438.

12. Seber GA, Lee AJ. Linear regression analysis. John Wiley & Sons 2012.

13. Tranmer M, Elliot M. Multiple linear regression. The Cathie Marsh Centre for Census and Survey Research (CCSR) 2008;5(5):1-5.

14. Padberg M. Linear optimization and extensions. Springer Science & Business Media 1995;12

15. Farrar DE, Glauber R R. Multicollinearity in regression analysis: the problem revisited. The Review of Economic and Statistics 1967;49:92-107.

16. https://www.kaggle.com/datasets/sazid28/advertising.csv