Exposure-Based Media Mix Modeling Using Machine Learning and Genetic Algorithms

An Audience Reach Optimization Framework for Advertising Budget Allocation

Thejan Dulara, Indra Mahakalanda, Prasanga Jayathunga
University of Moratuwa, Katubedda 10400, Sri Lanka

Abstract—**Media budget allocation remains a persistent challenge in the advertising industry. Inefficient spending and biased planning decisions often reduce campaign effectiveness. Advertisers struggle to balance investments across television, radio, press, and digital platforms while managing diminishing returns. This study proposes a data-driven media mix determination model that integrates supervised machine learning with genetic algorithm-based optimization. The objective is to maximize audience reach while maintaining cost efficiency. Unlike traditional media mix models that rely on aggregated medium-level performance, this study adopts an audience exposure-based modelling approach. Facebook and YouTube are used as digital media platforms in this study. Television and digital models are trained using exposure-based reach measures, such as 1 plus and 2 plus reach. Machine learning models, including decision trees, random forests, XGBoost, and LightGBM, are evaluated to capture complex and nonlinear relationships between spend and exposure-based reach. Smoothed reach response curves are used to identify efficiency levels and saturation points for each medium. A genetic algorithm is then applied to derive the optimal budget allocation across media under efficiency, reach, and cost constraints. The model is trained using real advertising data from the Sri Lankan market, ensuring practical relevance and applicability. Although the analysis is based on a country-specific dataset, the model is transferable to markets of similar scale. This study contributes to the literature by introducing an exposure-driven media mix modelling approach that improves media budget planning accuracy and supports more effective advertising decision-making.**

Keywords—Media mix determination; saturation points; audience reach; audience exposure

I. INTRODUCTION

Advertisers seek to reach the maximum possible audience while using their budgets efficiently [1]. However, allocating budgets across television, radio, print, and digital media remains a persistent challenge [1]. Each medium exhibits different response behaviors, and excessive spending on a single channel often results in diminishing returns [2]. These challenges increase the risk of inefficient budget allocation and reduced campaign effectiveness.

A data-driven approach that combines machine learning based prediction and optimization techniques can help address this problem [3]. Predictive models enable advertisers to understand how audience reach responds to varying spend levels across media. Optimization techniques then support the identification of an optimal media mix that maximizes reach at the lowest possible cost [3]. Media Mix Modeling (MMM) provides a structured framework for this purpose. Existing MMM approaches generally follow two directions. One focuses on quantitative objectives, such as maximizing reach while minimizing cost [4]. The other incorporates qualitative considerations, including campaign fitness, creative alignment, and historical effectiveness [4].

Most traditional MMM studies rely on aggregated spend to reach relationships at the media channel level [5]. Such approaches may overlook the effects of repeated audience exposure and the nonlinear nature of media response [6]. To address this limitation, this study adopts an exposure-based perspective. Television and digital media performance are modeled using exposure-based reach measures, such as 1 plus and higher reach levels, which reflect the proportion of the audience exposed to advertising multiple times [7]. This allows the model to learn directly from audience exposure behavior rather than general media-level aggregates.

Accordingly, this study develops a media mix determination model that integrates supervised machine learning techniques with genetic algorithm-based optimization. The primary focus is on quantitative performance factors. Unlike traditional aggregated media mix models that rely mainly on total spend to outcome relationships, the proposed framework models exposure-based reach metrics at different frequency thresholds. It incorporates measures such as 1 plus and higher exposure levels to capture audience response more precisely. This approach allows the model to reflect nonlinear response behavior and diminishing marginal effectiveness in a more direct manner. To the best of our knowledge, this Exposure-Based Media Mix Modelling framework, which combines exposure level modeling with machine learning and genetic algorithm optimization, represents a novel contribution to media planning research.

The remainder of this study discusses related concepts, defines the research problem and objectives, presents the methodology and results, and concludes with a discussion of implications.

II. RELATED CONCEPTS

A. Media Mix Modeling

Media Mix Modelling (MMM) is a widely used analytical approach that supports advertisers in allocating budgets across multiple media channels to improve sales performance and return on investment [8]. It enables the assessment of how individual media channels contribute to overall campaign outcomes within an evolving media environment [8]. Despite its usefulness, advertisers face several challenges when applying MMM, including limited data availability, interactions among media channels, and unpredictable consumer behaviour [9] [8]. To address these challenges, machine learning models such as Random Forest and gradient boosting techniques are increasingly applied. These models can capture complex response patterns by accounting for nonlinear effects, seasonality, and external economic factors [10]. As a result, MMM has become a critical tool for data-driven and cost-effective advertising decision-making [8].

MMM approaches can broadly be categorized into regression-based, Bayesian, and machine learning based frameworks. Traditional regression-based MMM typically relies on parametric specifications that incorporate advertising characteristics such as carryover and diminishing returns through transformations like ad stock and Hill functions [11]. These models offer clear interpretability, as regression coefficients can be directly translated into attribution metrics such as Return on Advertising Spend. However, their performance depends on correct functional form specification and sufficient variation in the underlying data [11].

Bayesian MMM extends regression frameworks by incorporating prior knowledge into the estimation process [12], [11]. Instead of producing single-point estimates, Bayesian approaches generate probability distributions for model parameters and attribution metrics, allowing uncertainty to be explicitly quantified [12]. This can be useful when data are sparse or highly correlated. In contrast, machine learning-based MMM methods emphasize predictive flexibility and nonlinear pattern recognition, often with fewer structural assumptions [10], [11]. While these approaches may reduce reliance on strict parametric forms, they may require additional techniques to ensure interpretability and stable attribution measurement [11].

B. The Concept of Diminishing Returns and Saturation Points in Media Advertising

An important concept in Media Mix Modeling is the principle of diminishing returns. Diminishing returns describe the nonlinear relationship between advertising spend and incremental audience response. As spending increases, each additional unit of investment generates progressively smaller gains in reach or engagement [13]. In empirical media response analysis, this behavior is commonly represented through saturation curves, such as S-curves or Hill transformations, which model how response flattens as spend rises [13].

For example, empirical observations in performance channels show that initial investment levels may produce high marginal returns, while additional increments of spend generate lower incremental gains. A retailer may observe a return on investment of 4:1 at moderate spend levels, declining to 2:1 or lower as spend increases further [13]. This decline reflects audience saturation, budget inefficiency, or limited incremental demand. The flattening of the response curve, therefore, represents the transition from efficient growth to diminishing effectiveness.

Saturation occurs when marginal return falls below a practical decision threshold or below the marginal return of alternative media channels [14]. In mathematical terms, the saturation point corresponds to the region where the first derivative of the response function approaches zero. At this stage, additional spending contributes negligible incremental reach or sales. Identifying this point is essential for rational budget allocation, as continued investment beyond saturation results in inefficient use of resources.

Different media channels exhibit distinct saturation behaviors. Television typically requires higher investment before saturation due to its broad reach potential, whereas digital platforms such as online video may reach saturation earlier because of targeting limits and audience overlap [2]. Moreover, saturation interacts with audience reach and frequency. When a campaign already reaches a high proportion of the addressable audience at multiple exposure levels, further impressions produce minimal incremental response [13].

Recognizing diminishing returns and saturation patterns is therefore fundamental to effective media planning. By estimating response curves and identifying efficiency thresholds, advertisers can avoid over investment in saturated channels and reallocate budgets toward media with remaining growth potential [14]. This empirical understanding forms the analytical foundation for exposure-based modeling and optimization in media mix determination.

C. Machine Learning Applications in Advertising

Machine learning techniques are well-suited for modeling the complex and nonlinear relationships commonly observed in advertising data. Models such as Random Forest, Support Vector Regression, and Bayesian Networks have been used to predict audience reach and identify saturation points across media channels [15]. These techniques also support adaptive decision-making by learning from historical performance data [15]. In addition, machine learning enables dynamic budgeting strategies, allowing advertisers to adjust media investments in response to observed performance changes, thereby improving efficiency and return on investment [16].

D. Audience Reach and Cost Efficiency

Audience reach and cost efficiency are fundamental metrics in advertising evaluation. Reach measures the proportion of the target audience exposed to advertising, while cost efficiency reflects how effectively budgets are converted into audience contact. Common indicators include Cost Per Click, return on investment, and incremental reach [17]. These metrics guide advertisers in reallocating budgets toward higher-performing channels. Prior research also highlights the efficiency gains achieved by shifting budgets from traditional media to digital platforms such as YouTube and display advertising networks, particularly when reach extension and targeting precision are prioritized [18].

E. Audience Exposure and Exposure Level-Based Planning

Traditional reach-based analysis often treats exposure as a binary outcome, without accounting for repeated audience contact. Exposure-based planning extends this concept by measuring reach at different exposure thresholds, such as 1 plus, 2 plus, or higher levels, which represent the proportion of the audience exposed to advertising at least a specified number of times [19]. Exposure frequency plays a critical role in advertising effectiveness, as repeated exposure influences recall, persuasion, and behavioral response. Modeling exposure levels allows advertisers to capture diminishing returns associated with repeated contact and to better understand audience response dynamics [20]. Exposure-based planning has been widely discussed in media planning literature and is commonly applied in television and digital advertising measurement frameworks. However, a detailed review of the theoretical literature on exposure frequency and marginal effectiveness is not included, as it is beyond the scope of this study.

F. Optimization Algorithms for Media Budget Allocation

Optimization techniques translate analytical insights into actionable media plans. Linear Programming and Genetic Algorithms are among the most commonly applied optimization methods in media budget allocation [21]. Linear Programming is effective for static problems with fixed constraints, whereas Genetic Algorithms offer greater flexibility for complex and nonlinear planning scenarios [21]. Genetic Algorithms can adapt to dynamic constraints, nonlinear media responses, and multiple objectives, making them well-suited for exposure-based media planning and budget optimization.

Genetic Algorithms are particularly useful in marketing environments characterized by large search spaces and nonlinear response behaviours [22]. Unlike deterministic allocation rules or simple heuristics, GA-based approaches iteratively evolve candidate solutions through selection, crossover, and mutation processes, allowing efficient exploration of multiple allocation combinations [22]. This makes them suitable for media planning problems where marginal returns differ across exposure levels, and constraints interact dynamically. Therefore, GA was selected as an appropriate optimization framework for handling the nonlinear and multi-constraint structure of the proposed exposure-based media mix model.

Together, these concepts form a comprehensive foundation for understanding and improving media mix determination using modern data-driven and optimization-based approaches.

III. Scope of the Study

This study addresses the problem of developing a media mix determination model for marketing campaigns that can predict optimized budget allocation while maximizing audience reach and maintaining cost efficiency. Advertisers face increasing difficulty in planning budgets across multiple media platforms due to nonlinear response behavior and diminishing returns at higher spend levels. This challenge becomes more complex when audience exposure effects are considered. Therefore, this study focuses on building a quantitative, data-driven framework that integrates exposure-based reach modeling with optimization techniques to support effective media planning decisions.

The objectives of this study are to identify key efficiency characteristics and optimal allocation strategies across media channels under different exposure levels. Specifically, the study aims to identify the maximum efficiency point and the zero efficiency point for each advertising medium across multiple exposure levels, including television, digital platforms, radio, and press. It further seeks to determine the reach and required budget corresponding to selected efficiency levels for each medium and exposure level. In addition, the study aims to estimate the budget and efficiency required to achieve a specified target reach under different exposure conditions. Another objective is to calculate the total achievable reach when all media operate at predetermined efficiency levels for selected exposure combinations. Finally, the study aims to optimize budget allocation across all media under predefined exposure levels in order to maximize total reach and identify the most efficient budget distribution for achieving target audience objectives.

IV. Methodology

A. Data

The data used in this study were provided by Third Shift Media (Pvt) Ltd, a leading advertising and media planning agency in Sri Lanka. The dataset consists of 18 separate datasets covering four main media channels: Television, Digital, Radio, and Press. Digital media data is limited to the Facebook and YouTube platforms. Within the television medium, ten separate datasets are provided. These datasets represent audience exposure levels from 1 plus to 10 plus reach. For Facebook and YouTube, exposure-based datasets corresponding to 1 plus, 4 plus, and 6 plus reach levels are available. These exposure thresholds were selected because they represent commonly applied frequency benchmarks in industry practice for digital campaign planning and performance evaluation. For radio and press, general audience exposure datasets are used without further exposure level separation.

Each dataset contains two variables: spend level and the corresponding audience exposure reach. Spend level represents the advertising expenditure for a given medium and is measured in Sri Lankan Rupees. Audience exposure reach represents the percentage of the total audience reached at a specific budget level for the corresponding exposure definition. This structure allows the analysis to capture how different levels of advertising exposure respond to increasing media spend across platforms.

B. Method of Analysis

The methodological framework of this study begins with a structured data preprocessing phase, as illustrated in Fig. 1. The raw datasets were examined for missing values, duplicate records, and outliers. No such issues were identified, and therefore, no additional data cleaning was required. Each dataset was then sorted by spend level to improve interpretability and ensure consistent model training.

A key preprocessing challenge was the inconsistency in population bases across media platforms. Television, radio, and press datasets were based on a population of 17.6 million. In contrast, Facebook exposure data were based on a population of 8.05 million, while YouTube exposure data reflected a population of 7.47 million. To ensure comparability across all media, Facebook and YouTube datasets were scaled proportionally to match the 17.6 million population base.

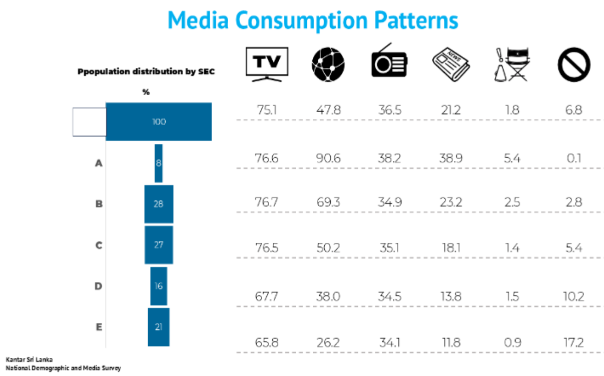

Fig. 1. Media consumption patterns (National Demographic and Media Survey by Kantar Sri Lanka).

The primary objective of the analysis was to identify efficiency levels and saturation points for each medium and exposure level. Since the relationship between advertising spend and audience exposure-based reach is nonlinear and exhibits diminishing returns, four supervised machine learning models were evaluated. These models were Decision Tree, Random Forest, LightGBM, and XGBoost. For each exposure-based dataset across all four media, all four models were trained independently. Model performance was evaluated using Mean Squared Error and R squared metrics on both training and testing sets. The model with the highest R squared value and lowest Mean Squared Error was selected for subsequent analysis.

To support high-resolution analysis, a continuous range of spend values was generated within the observed minimum and maximum spend levels of each dataset. Predicted exposure-based reach values were obtained across this range using the selected model, forming a dense response curve. Since raw predictions could contain minor fluctuations, a Gaussian smoothing filter was applied. This ensured stable derivative estimation in later steps.

The first Numerical derivatives of the smoothed reach curves were computed to quantify marginal reach per unit spend, defined as efficiency. Efficiency values were normalized to a percentage scale, where 100 percent represented the point of maximum marginal reach. Two key thresholds were identified from each efficiency curve. The first was the maximum efficiency point. The second was the saturation point, defined as the point where efficiency fell below 2 per cent of its maximum value, indicating negligible incremental reach.

These efficiency curves enabled two complementary budgeting strategies. In the efficiency-based approach, a user-defined efficiency threshold, such as 80 per cent was selected, and the corresponding budget and reach were identified after the maximum efficiency point. In the reach-based approach, a target reach level was specified, and the corresponding budget and efficiency were derived from the prediction curve.

This process was repeated for all 18 exposure-based datasets across the four media channels. The next phase involved estimating the combined reach under a multi-channel media planning scenario. Media layering order was determined using the National Demographic and Media Survey by Kantar Sri Lanka (Fig. 2). Based on audience penetration levels, media were prioritized as television first, followed by digital platforms in the order of Facebook and YouTube, then radio, and finally press.

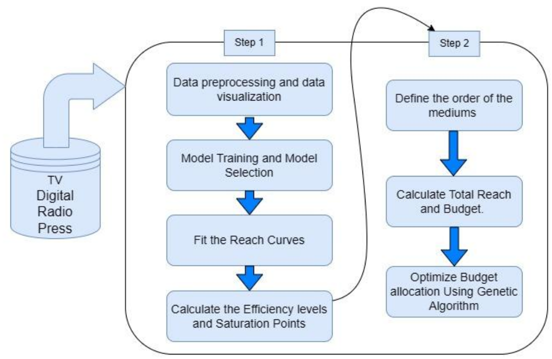

Fig. 2. Research process.

Direct aggregation of reach across channels would lead to duplication errors due to overlapping audiences. To address this issue, realistic duplication assumptions were introduced. It was assumed that 60 per cent of the Facebook audience overlapped with television viewers. 90 per cent of the YouTube audience overlapped with audiences reached through television and Facebook. Similarly, 90 per cent of the radio audience overlapped with audiences reached through television and digital platforms. Finally, 95 per cent of the press audience overlapped with all other media. These assumptions were based on observed media consumption patterns.

To incorporate duplication effects, digital, radio, and press models for all exposure levels were retrained using adjusted datasets that excluded overlapping audiences. Television models were used to estimate the initial reach, followed sequentially by Facebook, YouTube, radio, and press. All modeling steps, including prediction, smoothing, derivative calculation, and efficiency-based analysis, were repeated using the adjusted datasets.

For each medium and exposure level, the budget and reach corresponding to a selected efficiency threshold were extracted. By summing results across media, the total budget required and cumulative reach were estimated. The model also allows users to compute achievable reach under different exposure combinations, such as television 1 plus with Facebook 4 plus and YouTube 4 plus, alongside radio and press at selected efficiency levels.

To further optimize budget allocation under practical constraints, a Genetic Algorithm was applied. The algorithm aimed to maximize total reach while minimizing total spend, or alternatively, achieve a target reach at the lowest possible cost. The initial population consisted of candidate budget plans across selected exposure levels of television, digital, radio, and press, subject to total budget constraints. Each plan was evaluated using trained media models to predict reach. Fitness scores incorporated penalties for exceeding budgets or failing to meet reach targets.

The algorithm evolved solutions over multiple generations, typically between 30 and 50. Selection retained high-performing plans, crossover combined strong solutions, and mutation introduced controlled variation. A repair mechanism ensured all solutions satisfied constraints. Initial populations were seeded using insights from efficiency analysis to improve convergence.

The final output included the optimal budget allocation, predicted reach for each medium, and corresponding efficiency levels. This enabled the Genetic Algorithm to solve different planning scenarios, including fixed budget allocation and target reach achievement.

V. Results

To predict exposure level-specific reach as a function of media spend, machine learning models were selected based on their predictive performance. Decision Tree models were selected for all television exposure level models and for radio, as these achieved the highest R squared values with low Mean Squared Errors. Random Forest models were selected for all digital exposure level models, including Facebook and YouTube, as well as for the press. These models consistently outperformed alternatives across exposure levels.

TABLE I. Results of Selected Machine Learning Models Before Addressing Audience Duplication.

Medium	Selected Model	Frequency	R2	MSE
TV	Decision Tree	1+	0.9853	1.0705
TV	Decision Tree	2+	0.9919	0.8514
TV	Decision Tree	3+	0.9969	0.3978
TV	Decision Tree	4+	0.9983	0.2606
TV	Decision Tree	5+	0.9986	0.2528
TV	Decision Tree	6+	0.9990	0.1913
TV	Decision Tree	7+	0.9992	0.1688
TV	Decision Tree	8+	0.9993	0.1500
TV	Decision Tree	9+	0.9994	0.1433
TV	Decision Tree	10+	0.9994	0.1442
Facebook	Random Forest	1+	0.9999	0.0004
Facebook	Random Forest	4+	0.9999	0.0028
Facebook	Random Forest	6+	0.9999	0.0015
YouTube	Random Forest	1+	0.9998	0.0235
YouTube	Random Forest	4+	0.9999	0.0120
YouTube	Random Forest	6+	0.9999	0.0097
Radio	Decision Tree	-	0.9964	0.1317
Press	Random Forest	-	0.9942	0.0301

Model performance was evaluated using R squared and Mean Squared Error metrics on both training and testing datasets. Results for all exposure level models before addressing audience duplication are summarized in Table I. Performance results after adjusting for duplication are presented for digital, radio, and press models in Table II. Since television served as the base medium, duplication adjustment was not required for television models. To assess potential overfitting, R squared values were calculated separately for training and testing datasets for each model, and the absolute difference between them was computed. Across all exposure level models, this difference remained below a threshold of 0.05. Therefore, no substantial divergence between training and testing performance was observed, indicating stable generalization and a limited risk of overfitting or underfitting in the analyzed advertising datasets.

TABLE II. Results of Selected Machine Learning Models After Addressing Audience Duplication.

Medium	Selected Model	Frequency	R2	MSE
Facebook	Random Forest	1+	0.9999	0.00007
Facebook	Random Forest	4+	0.9999	0.0004
Facebook	Random Forest	6+	0.9999	0.00024
YouTube	Random Forest	1+	0.9998	0.0002
YouTube	Random Forest	4+	0.9999	0.0001
YouTube	Random Forest	6+	0.9999	0.00009
Radio	Decision Tree	-	0.9964	0.0013
Press	Random Forest	-	0.9942	0.0001

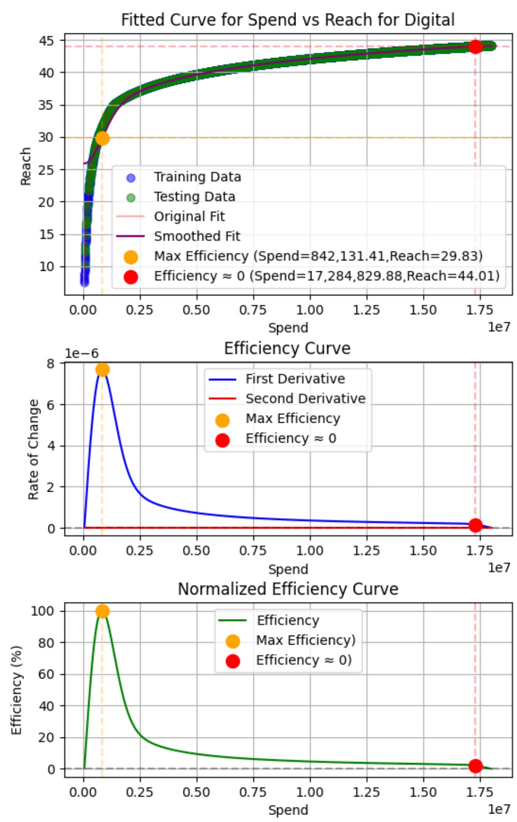

Fig. 3. Reach curve, efficiency curve, and normalized efficiency curve for the Facebook medium at the 1+ exposure level before duplication adjustment.

Based on the selected models, reach response curves were generated for all media and exposure levels under both duplication-adjusted and non-adjusted conditions. For each model, reach curves, efficiency curves, and normalized efficiency curves were derived. Fig. 3 illustrates the model outputs for a representative Facebook digital media case at the 1+ exposure level before addressing duplication. Fig. 4 presents the reach curve after duplication adjustment. The plots highlight the points of maximum efficiency and saturation.

Budget and reach outcomes obtained using both efficiency-based and reach-based targeting approaches, before and after addressing duplication scenarios, are summarized across all exposure levels and media in Table III to Table XX.

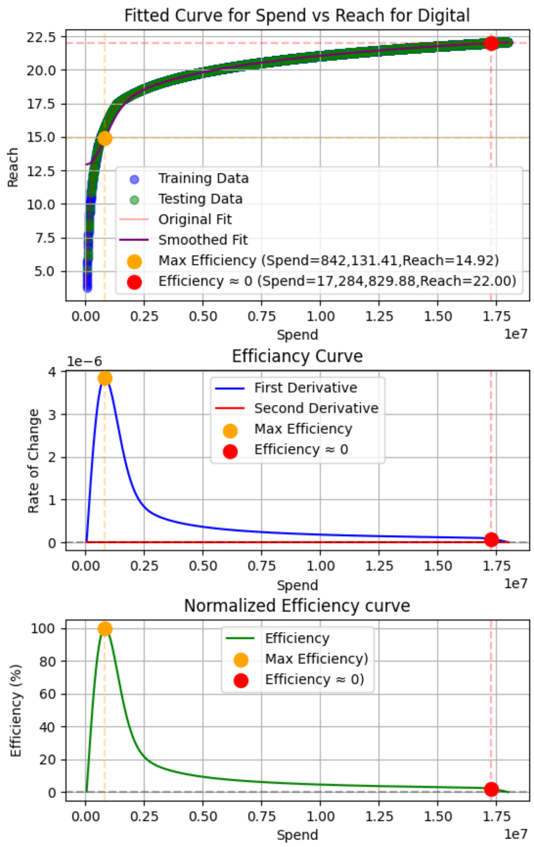

Fig. 4. Reach curve, efficiency curve, and normalized efficiency curve for the Facebook medium at the 1+ exposure level after duplication adjustment.

TABLE III. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE TV 1+ EXPOSURE LEVEL.

Efficiency	Spend (LKR - Mn)	Reach % (Before)
0%	125	74.38
20%	21.90	65.74
40%	14.55	62.53
60%	11.52	60.19
80%	9.35	57.83
100%	6.14	53.18

TABLE IV. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE TV 2+ EXPOSURE LEVEL.

Efficiency	Spend (LKR - Mn)	Reach % (Before)
0%	125.00	74.14
20%	20.96	63.95
40%	14.51	60.20
60%	11.60	57.24
80%	9.45	54.14
100%	6.22	47.96

TABLE V. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE TV 3+ EXPOSURE LEVEL.

Efficiency	Spend (LKR - Mn)	Reach % (Before)
0%	125.00	73.67
20%	20.87	62.10
40%	14.80	57.32
60%	11.89	54.32
80%	9.68	50.50
100%	6.35	42.90

TABLE VI. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE TV 4+ EXPOSURE LEVEL.

Efficiency	Spend (LKR - Mn)	Reach % (Before)
0%	125.00	73.25
20%	21.42	60.67
40%	15.22	55.83
60%	12.20	51.70
80%	9.91	47.26
100%	6.49	38.49

TABLE VII. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE TV 5+ EXPOSURE LEVEL.

Efficiency	Spend (LKR - Mn)	Reach % (Before)
0%	125.00	79.98
20%	23.01	60.01
40%	15.88	54.15
60%	12.69	49.52
80%	10.27	44.55
100%	6.69	34.79

TABLE VIII. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE TV 6+ EXPOSURE LEVEL.

Efficiency	Spend (LKR - Mn)	Reach % (Before)
0%	125.00	72.58
20%	24.00	58.85
40%	16.59	52.66
60%	13.31	47.78
80%	10.77	42.44
100%	6.98	31.87

TABLE IX. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE TV 7+ EXPOSURE LEVEL.

Efficiency	Spend (LKR - Mn)	Reach % (Before)
0%	125.00	72.13
20%	25.664	58.21
40%	17.330	51.26
60%	13.90	46.14
80%	11.24	40.49
100%	7.26	29.31

TABLE X. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE TV 8+ EXPOSURE LEVEL.

Efficiency	Spend (LKR - Mn)	Reach % (Before)
0%	125.00	71.77
20%	27.31	57.74
40%	18.13	50.20
60%	14.59	44.96
80%	11.79	39.10
100%	7.57	27.41

TABLE XI. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE TV 9+ EXPOSURE LEVEL.

Efficiency	Spend (LKR - Mn)	Reach % (Before)
0%	125.00	71.31
20%	29.12	57.38
40%	19.22	49.34
60%	15.42	43.89
80%	12.49	37.77
100%	7.92	25.62

TABLE XII. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE TV 10+ EXPOSURE LEVEL.

Efficiency	Spend (LKR - Mn)	Reach % (Before)
0%	125.00	70.95
20%	30.61	56.87
40%	20.14	48.52
60%	16.16	42.95
80%	13.05	36.76
100%	8.31	24.29

TABLE XIII. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE FACEBOOK 1+ EXPOSURE LEVEL BEFORE AND AFTER DUPLICATION SCENARIOS.

Efficiency	Spend (LKR - Mn)	Reach % (Before	Reach % (After)
0%	17.99	44.07	22.03
20%	2.62	37.23	18.62
40%	1.87	35.64	17.82
60%	1.54	34.37	17.19
80%	1.27	32.93	16.46
100%	0.84	29.83	14.92

TABLE XIV. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE FACEBOOK 4+ EXPOSURE LEVEL BEFORE AND AFTER DUPLICATION SCENARIOS.

Efficiency	Spend (LKR - Mn)	Reach % (Before	Reach % (After)
0%	1.32	29.5	11.80
20%	1.22	29.15	11.66
40%	1.03	27.32	10.93
60%	0.86	24.45	9.78
80%	0.72	21.5	8.6
100%	0.28	8.67	3.47

TABLE XV. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE FACEBOOK 6+ EXPOSURE LEVEL BEFORE AND AFTER DUPLICATION SCENARIOS.

Efficiency	Spend (LKR - Mn)	Reach % (Before	Reach % (After)
0%	1.68	28.05	11.22
20%	1.63	27.97	11.19
40%	1.53	27.22	10.89
60%	1.32	24.95	9.98
80%	1.09	21.72	8.69
100%	0.42	8.77	3.51

TABLE XVI. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE YOUTUBE 1+ EXPOSURE LEVEL BEFORE AND AFTER DUPLICATION SCENARIOS.

Efficiency	Spend (LKR - Mn)	Reach % (Before	Reach % (After)
0%	23.7	23.3	2.33
20%	6.02	18.3	1.83
40%	3.31	16.6	1.66
60%	2.65	15.8	1.58
80%	2.18	14.9	1.49
100%	1.49	13.3	1.33

TABLE XVII. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE YOUTUBE 4+ EXPOSURE LEVEL BEFORE AND AFTER DUPLICATION SCENARIOS.

Efficiency	Spend (LKR - Mn)	Reach % (Before	Reach % (After)
0%	23.7	18.3	1.83
20%	5.74	13.7	1.37
40%	3.81	11.8	1.18
60%	3.1	10.5	1.05
80%	2.54	9.1	0.91
100%	1.7	6.2	0.62

TABLE XVIII. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES AT THE YOUTUBE 6+ EXPOSURE LEVEL BEFORE AND AFTER DUPLICATION SCENARIOS.

Efficiency	Spend (LKR - Mn)	Reach % (Before	Reach % (After)
0%	23.7	17.3	1.73
20%	7.61	13.3	1.33
40%	5.1	11.1	1.11
60%	3.96	9.3	0.93
80%	3.16	7.5	0.75
100%	2.12	4.5	0.45

TABLE XIX. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES FOR RADIO BEFORE AND AFTER DUPLICATION SCENARIOS.

Efficiency	Spend (LKR - Mn)	Reach % (Before	Reach % (After)
0%	10.00	39.02	3.9
20%	1.95	30.94	3.09
40%	1.13	28.02	2.8
60%	0.89	26.47	2.65
80%	0.73	24.95	2.5
100%	0.48	21.94	2.19

TABLE XX. BUDGET AND REACH FOR EFFICIENCY-BASED AND REACH-BASED TARGETING APPROACHES FOR PRESS BEFORE AND AFTER DUPLICATION SCENARIOS.

Efficiency	Spend (LKR - Mn)	Reach % (Before	Reach % (After)
0%	13.1	27.46	1.37
20%	1.61	23.22	1.16
40%	1.25	22.39	1.12
60%	1.05	21.58	1.08
80%	0.88	20.58	1.03
100%	0.59	18.37	0.92

Using the efficiency curves, total achievable reach and required budget were computed for selected efficiency thresholds. Table XXI presents a planning scenario with efficiency targets of 40 per cent for television 1 plus exposure, 45 per cent for Facebook 4 plus exposure, 50 per cent for YouTube 4 plus exposure, 60 per cent for radio, and 70 per cent for press. These results demonstrate how exposure level selection and efficiency targets influence total reach and budget requirements.

TABLE XXI. COMBINED BUDGET AND REACH OUTCOMES

Model	Efficiency	Spend (LKR)	Reach (%)
TV 1+	40%	14,556,393.14	62.53
Facebook 4+	45%	1,006,975.7	10.76
YouTube 4+	50%	3,427,182.72	1.12
Radio	60%	896,829.68	2.65
Press	70%	974,032.4	1.06
Total		**20,861,413.64**	**78.12**

To further enhance budget allocation decisions, a Genetic Algorithm was applied. In the first scenario, the algorithm optimized a fixed total budget of LKR 20 million across television 1 plus, Facebook 4 plus, YouTube 4 plus, radio, and press. The resulting optimal allocation and predicted reach are reported in Table XXII. In the second scenario, the algorithm was configured to achieve a target reach of 85 per cent for the same exposure level combination. The resulting budget allocation and medium-wise reach outcomes are presented in Table XXIII. These results confirm the effectiveness of the proposed model in supporting exposure-based media planning and optimization.

TABLE XXII. GA RESULTS FOR TARGET BUDGET OF LKR 20 MILLION SPEND.

Model	Efficiency	Spend (LKR)	Reach
TV 1+	16.88%	12,509,660.63	62.049%
Facebook 4+	2.68%	1,347,501.75	11.877%
YouTube 4+	52.21%	2,702,591.28	1.009
Radio	5.97%	2,563,603.05	3.239%
Press	25.71%	876,643.29	1.103
Total		**20,000,000.00**	**79.278%**

TABLE XXIII. GA RESULTS FOR TARGET REACH OF 85% OF TOTAL AUDIENCE.

Model	Efficiency	Spend (LKR)	Reach
TV 1+	4.8%	27,901,264.60	67.52%
Facebook 4+	6.5%	1,310,219.54	11.84%
YouTube 4+	27.4%	3,702,008.25	1.19%
Radio	5.2%	3,036,680.57	3.32%
Press	12.8%	1,080,621.53	1.13%
Total		**37,030,794.49**	**85.00%**

VI. DISCUSSION

This study aimed to develop a practical and data-driven media mix determination model suitable for real-world advertising decision-making using Sri Lankan media data. The proposed framework supports key planning tasks by identifying performance patterns, efficiency levels, and optimal budget allocations across media channels. By incorporating exposure-based reach measures, the model extends traditional media mix modeling approaches and provides a more realistic representation of audience response behavior.

In line with the study objectives, the analysis identified the maximum efficiency points and saturation points for television, digital platforms including Facebook and YouTube, radio, and press across different exposure levels. These results provide valuable insights for advertisers by highlighting spend regions that yield high marginal returns and those that result in inefficient spending. The study further addressed efficiency-based and reach-based planning objectives by estimating the budget required to achieve specified efficiency levels and target reach values under different exposure conditions. This enables advertisers to answer practical planning questions related to expected reach outcomes and budget requirements.

The study also demonstrated how total achievable reach can be estimated by operating all media at predefined efficiency levels for selected exposure combinations. This provides advertisers with clear guidance on how exposure level selection influences cumulative reach outcomes. Budget optimization objectives were addressed through the application of a Genetic Algorithm, which identified optimal budget allocations under constraints such as fixed budgets and target reach levels. The optimization results show that exposure-aware planning can improve budget efficiency and overall campaign reach.

Although the analysis is based on industry data from Sri Lanka, the model is not restricted to a single media plan or organization. The framework can be adapted for advertising budget allocation decision support in small and medium-scale companies as well as larger conglomerates operating in similar market contexts. Overall, the proposed approach offers a simple yet robust tool for exposure-driven media planning and informed budget optimization.

VII. Conclusion and Implications

This study presents a practical and data-driven framework for media mix determination that integrates exposure-based reach modeling with machine learning and genetic algorithm optimization. By focusing on audience exposure levels rather than aggregated channel performance, the proposed approach provides more realistic insights into media response behavior and budget efficiency. The model supports advertisers in identifying efficient spending zones, estimating achievable reach under different exposure conditions, and optimizing budget allocation across multiple media platforms.

Despite its practical contributions, the study has several limitations. The proposed framework focuses exclusively on quantitative performance measures and does not incorporate qualitative factors such as campaign fitness, brand positioning, creative effectiveness, or prevailing market trends. Therefore, advertisers should use the model as a decision support tool rather than a standalone planning solution. Strategic judgment and contextual knowledge remain essential when finalizing media plans.

The model was developed using a duplication-adjusted audience scenario based on observed media consumption patterns. However, audience overlap varies depending on campaign objectives, target segments, and media sequencing strategies. As a result, the duplication parameters applied in this study may not fully reflect all real-world scenarios. In addition, the statistical stability of the results depends on the quality, consistency, and historical depth of the advertising datasets used. The findings may therefore be sensitive to data limitations, exposure definitions, and duplication adjustments implemented during the modeling process.

Future research may integrate qualitative considerations through structured multi-criteria decision-making approaches such as the Analytic Hierarchy Process (AHP). An AHP-based extension could incorporate expert evaluations of qualitative campaign attributes and combine them with the quantitative optimization outputs. This would allow media planning decisions to reflect both measurable performance indicators and strategic brand-level considerations, resulting in a more comprehensive decision framework.

Finally, the model framework is designed in a way that it can be retrained using industry-specific or campaign-specific data, although external validation across different markets has not been conducted within this study. This allows the framework to be adapted for different markets, exposure definitions, and planning objectives. Future extensions may integrate qualitative indicators or dynamic market signals to further enhance decision-making accuracy and practical relevance.

Acknowledgment

The authors would like to acknowledge Third Shift Media (Pvt) Ltd for providing industry data, domain knowledge, and professional support that contributed to this study. The authors also acknowledge the University of Moratuwa for the academic guidance and research environment. Co-authors are gratefully acknowledged for their contributions to this work. In addition, the authors acknowledge the use of ChatGPT version 5.2 for language refinement and clarity improvement during manuscript preparation.

References

[1] "Key Insights from the WFA 2025 Media Budgets Survey," https://www.abintus.consulting/blog/key-insights-from-the-wfa-2025-media-budgets-survey

[2] "Profit Ability 2: The new business case for advertising," https://www.thinkbox.tv/research/thinkbox-research/profit-ability-2-the-new-business-case-for-advertising

[3] M. True, "Optimizing ad spend with AI-powered marketing mix modeling," https://www.fastcompany.com/91250176/optimizing-ad-spend-with-ai-powered-marketing-mix-modeling

[4] Y. Fareniuk and G. Chornous, "Optimization of Media Strategy via Marketing Mix Modeling in Retailing. ," https://www.journals.vu.lt/ekonomika/article/view/27502/30587

[5] J. Browining, "What is Media Mix Modeling (MMM)?," https://www.northbeam.io/blog/media-mix-modeling-mmm-guide

[6] R. Dew, N. Padilla, and A. Shchetkina, "Your MMM is Broken: Identification of Nonlinear and Time-varying Effects in Marketing Mix Models," Aug. 2024

[7] V. R. Lane, "Frequency (Marketing)," https://en.wikipedia.org/wiki/Frequency_%28marketing%29?utm , pp. 80–91, doi: 10.1509/jmkg.64.2.80.17996

[8] Y. Liu, J. Laguna, M. Wright, and H. He, "Media mix modeling – A Monte Carlo simulation study," Journal of Marketing Analytics, vol. 2, no. 3, pp. 173–186, Sep. 2014, doi: 10.1057/jma.2014.3

[9] D. Chan and M. Perry, "Challenges and Opportunities in Media Mix Modeling," https://www.semanticscholar.org/paper/Challenges-and-Opportunities-in-Media-Mix-Modeling-Chan-Perry/c53c3ff6241c31a05d16415c014d4f9ce2cc9d29

[10] S. Kisilevich, "Improving Marketing Mix Modeling Using Machine Learning Approaches | Towards Data Science," https://towardsdatascience.com/improving-marketing-mix-modeling-using-machine-learning-approaches-25ea4cd6994b/

[11] L. Scharfe, J. Antje, and A. Malcherek, "Regression, Bayesian and Machine Learning Methods for Media Mix Modeling-A Comparative Study Motivation and Research Objectives."

[12] K. D. Systems, "Why Bayesian Marketing Mix is Superior to Traditional Approaches?," https://keends.com/blog/why-is-bayesian-regression-modeling-approach-superior-to-traditional-methods/

[13] "Diminishing returns in marketing: when your next euro delivers less than your last - Analytical Alley," https://www.analyticalalley.com/knowledge-hub/diminishing-returns-curve-marketing

[14] M. Lascano, "Saturation Curves in MMM - Indaru," https://www.indaru.com/mmm-saturation-curves/

[15] Y. Li, "Advertising Optimization and Feature Analysis Based on Machine Learning," INSTICC, Aug. 2024, pp. 70–76. doi: 10.5220/0012822000004547

[16] J. A. Choi and K. Lim, "Identifying machine learning techniques for classification of target advertising," Sep. 2020, Korean Institute of Communications Information Sciences. doi: 10.1016/j.icte.2020.04.012

[17] K. Ganesh, C. Thanu Sri, K. Purnendra Reddy, H. Niyotwizeye Gisele, and C. Gloria, "Effectiveness of Online Advertising in Reaching Target Audiences," 2024. [Online]. Available: www.multiresearchjournal.com

[18] Y. Jin, S. Shobowale, J. Koehler, and H. Case, "The Incremental Reach and Cost Efficiency of Online Video Ads over TV Ads."

[19] "Reach and frequency," https://www.thinkbox.tv/how-to-use-tv/spots/reach-and-frequency

[20] G. Broussard, "How advertising frequency can work to build online advertising effectiveness," International Journal of Market Research, vol. 42, pp. 439–457, 2000, doi: 10.1177/147078530004200406

[21] "Webinar: Media Mix Modeling and Optimization," https://www.bayesia.com/bayesialab/tutorials/Webinar-Media-Mix-Modeling-and-optimization

[22] "Genetic Matching Algorithms: Genetic Algorithms in Marketing: Crafting the Perfect Ad Campaign - FasterCapital," https://fastercapital.com/content/Genetic-Matching-Algorithms--Genetic-Algorithms-in-Marketing--Crafting-the-Perfect-Ad-Campaign.html#Success-Stories-in-Genetic-Algorithm-Marketing