Metaheuristic Optimization Review

♦ ISSN: 3066-280X Vol: 06, No. 01, 2026 ♦ PP. 15–28
Article DOI: https://doi.org/10.54216/MOR.060102

RESEARCH ARTICLE

Optimizing Digital Marketing Revenue Forecasting Using an XGBoost–Dipper Throated Optimization Hybrid Model

Mohamed Rabehi[1] **Abdelaziz Rabehi**[2,*]

[1] Laboratory Department of Civil Engineering, University of Djelfa, 17000 Djelfa, Algeria
[2] Telecommunications and Smart Systems Laboratory, University of Djelfa, PO Box 3117, Djelfa 17000, Algeria

Emails: Mohamed.Rabeh@gmail.com · Abdelaziz.rabehi@univ-djelfa.dz

Received: December 10, 2025 Revised: February 05, 2026 Accepted: April 04, 2026 ⋆ Corresponding author

ABSTRACT

The explosive growth of digital marketing data and the increasing need for accurate revenue forecasting have driven the adoption of advanced Machine Learning (ML) techniques capable of modeling complex, nonlinear relationships in dynamic environments. Motivated by the limitations of traditional linear forecasting methods, this study proposes an optimized predictive framework that integrates the Extreme Gradient Boosting (XGBoost) algorithm with a novel metaheuristic, Dipper Throated Optimization (DTO), to enhance model performance on temporal marketing data. The key contribution of this work lies in combining ensemble learning with bio-inspired optimization to achieve superior predictive accuracy and stability in Time-Series forecasting tasks. As the experiments of the Digital Marketing Metrics dataset demonstrate, the original XGBoost model achieved a Mean Squared Error (MSE) of 0.0905 and a coefficient of determination (R^2) of 0.8007, and the optimized XGBoost+DTO model has significantly improved results, with an MSE of 0.0010 and a coefficient of determination (R^2) of 0.9002. These results support the argument that DTO is effective in hyperparameter optimization and reducing generalization errors. The results of this paper are not unique to digital marketing, and the authors have presented a scalable, interpretable optimization model that can be generalized to other data-intensive fields, such as financial analytics, demand forecasting, and customer behavior modelling. The study is a good step in the right direction of creating more accurate, adaptive and data-driven decision-making in the digital economy by integrating ML and nature-inspired optimization.

Keywords: Digital Marketing Analytics ▪ Machine Learning (ML) ▪ Extreme Gradient Boosting (XGBoost) ▪ Dipper Throated Optimization (DTO) ▪ Revenue Forecasting

1. INTRODUCTION

The fact is that the development and introduction of information technology in the world economy have driven unprecedented growth in information creation, especially in digital marketing. Companies across all sectors are turning to digital platforms to reach their customers, customize their campaigns, and gauge their marketing effectiveness. This has led businesses to face large volumes of non-homogeneous data resulting from websites, social media, email campaigns, and internet transactions. The problem, however, lies not only in gathering such data but also in efficiently analyzing it to produce actionable insights that improve marketing performance and increase profitability. In this regard, accurate revenue forecasting has emerged as a burning, though continuous, issue. E-business organizations widely use promotional campaigns to reach specific consumer groups, improve consumer engagement, and boost sales [1]. Nevertheless, revenue performance is not well-predictable due to the dynamism of digital markets and multifactorial relationships among consumers, campaigns, and competitive environments.

Conventional forecasting techniques are inadequate for dealing with the complex relationships in marketing data. Nonlinear, interdependent characteristics that affect revenue, such as consumer behavior, campaign timing, product seasonality and competitor activity, are often interdependent. Linear statistical models, despite their usefulness in certain situations, struggle to model nonlinearities and unobservable dependencies [2]. This means that companies face challenges in making reliable forecasts, leading to the wasteful use of marketing funds and missed opportunities to optimize. This is a gap that needs more advanced analytical methods that can account for complex interactions and display latent structures in the data. To counter these issues, it is crucial to employ analytical paradigms that extend beyond traditional regression analysis and instead leverage more sophisticated computational intelligence tools that can effectively handle data heterogeneity, very high dimensionality, and nonlinearity .

Machine Learning (ML) and Artificial Intelligence (AI) are now groundbreaking technologies; in this respect, they are one of the technologies in the field of marketing analytics. This is due to their inherent capabilities for working with large volumes of data, finding latent relationships, and performing predictive analytics, which make them viable for forecasting revenues. Machine learning algorithms can learn on their own, accommodate novel patterns, and become more accurate over time [3]. ML and AI models are more adaptable than traditional statistical models, which do not require linearity or independence assumptions. Such flexibility is well-suited to digital marketing, where data are often high-dimensional, dynamic, and contextual, such as consumer mood, brand attitude, and macroeconomic factors.

The use of AI and ML in marketing is not confined to prediction; they can also provide insights into what drives revenue and customer response. For example, evaluating feature importance in models such as XGBoost or Random Forests can identify which marketing communications, demographic factors, or other engagement metrics play the most significant role in final revenue results [4]. This explanatory power can help organizations not only make more accurate predictions but also make better strategic decisions, whether in budget allocation or product positioning. Consequently, AI and ML do not remain merely tools of analysis but are involved in the strategic management of the marketing process and become its constituents [5]. An adoption will enable organizations to be more productive, closer to customers, and acquire a sustainable competitive advantage in the digital realm.

The data available and its quality are the core of any AI or ML model's success. Predictive modeling relies on precise, thorough databases that accurately reflect the phenomena they represent. Strict data preprocessing and analysis is then the first and most crucial step towards developing a trustworthy machine learning model [6]. In this work, the Digital Marketing Metrics set serves as the baseline data, containing the most significant indicators of marketing performance and their associations with revenue. Nonetheless, problems in the raw marketing data include noise, missing values, redundant attributes, and irrelevant features, which can adversely affect model performance and reliability. To solve these issues, the primary steps in effective data preprocessing include data cleaning, feature selection, dimensionality reduction, and scaling. Feature selection techniques are used to eliminate irrelevant or redundant variables, thereby simplifying the models provided. Normalization also ensures that the input features make the same contribution to the learning process. Nevertheless, even with ready-made data, to develop a really effective predictive model, attention should be paid to parameter tuning. There can also be many hyperparameters in machine learning algorithms that control learning rates, tree depth, and regularization coefficients, which can significantly influence the model's performance. The hyperparameters are not easily optimized, especially in non-convex and high-dimensional search spaces. Conventional methods, e.g., grid or random search, can be computationally expensive and may not yield a globally optimal solution. To solve these problems, metaheuristic optimization methods have gained popularity because they are flexible and can serve as powerful alternatives . These algorithms are inspired by the natural, biological, or physical sciences to explore and exploit the search space.

Among the vast number of metaheuristic algorithms, the Dipper Throated Optimization (DTO) algorithm has recently attracted attention for its ability to solve highly complex optimization problems [7]. DTO mimics the hunting and movement behaviour of the dipper bird, which is dynamic, allowing the bird to be innovative and flexible as it navigates the solution space. This ability to balance exploration (discovering new things) and exploitation (leveraging existing solutions) is beneficial for tuning machine learning models. Combining DTOs with predictive models like XGBoost is a powerful strategy for improving performance, minimizing errors, and strengthening predictions.

XGBoost (Extreme Gradient Boosting) is one of the best and most widely used ensemble learning algorithms for structured data. It builds a forest of decision trees, step by step, in which each successive tree is intended to reduce the errors of the other trees and thereby minimize the loss function using gradient boosting trees . The advantages of XGBoost include scalability, regularization to prevent overfitting, and the ability to handle sparse data. However, tuning the maximum depth, subsample ratio, and learning rate is essential for achieving maximum performance. In this paper, a hybrid model will be created that uses the DTO algorithm to adjust hyperparameters, enhancing XGBoost's predictive capabilities, and, at the same time, improving DTO's adaptive search, thereby boosting accuracy and reducing computation time.

2. LITERATURE REVIEW

The use of Data Sciences, which enhance decision-making and facilitate the extraction of actionable insights from large datasets in the digital marketing landscape, has grown significantly over the past decade. In the analysis provided by [8], it is evident that despite these advancements, there is still limited evidence on strategies to optimize the management of Data Sciences, including machine learning, within digital marketing. This study, therefore, aims to review the analytical methods, practical applications, and performance metrics related to Data Sciences as applied to digital marketing techniques and strategies. The comprehensive literature review uncovers significant applications and contributes to understanding how innovative Data Mining and knowledge discovery techniques can be developed and implemented effectively. Important theoretical implications are highlighted, providing a foundation for subsequent studies, and recommendations are made to guide businesses, marketers, and non-technical researchers in designing robust digital marketing strategies.

Machine learning (ML) holds significant potential for various marketing applications. Following the work of [9], the diverse range of data types, methodologies, tools, and programming languages often complicates the integration of knowledge within marketing analytics teams, hindering effective collaboration. Visual-based programming emerges as a promising solution to these challenges by enabling the more intuitive orchestration of ML projects through graphical interfaces. One noteworthy tool highlighted for its capabilities in this area is the KNIME Analytics Platform, recognized for simplifying the development and execution of ML workflows. This software supports a collaborative environment where marketing professionals can learn, share, and reuse workflows via a dynamic repository hosted on the KNIME hub. To contribute to the practical application of ML in marketing, the authors curated five annotated projects focusing on key areas such as customer churn prediction, sentiment analysis, automated image analysis, search engine optimization, and enhancing customer experience. In their work, two of these projects are detailed with comprehensive step-by-step guides to facilitate learning for practitioners, academics, and ML enthusiasts.

In online marketing, experimenting with new website content is a crucial tactic for increasing customer engagement. As stated in [10], creating effective marketing content remains relatively manual and time-consuming, with no clear guidelines for marketers to follow. To address the gap between content development and online testing, the present study provides AI-driven, practical insights derived from historical data to improve the creative process. The authors introduce a neural-network-based system that is meant to analyze marketing content and extract insights from it. Namely, a multimodal neural network is applied to predict the attractiveness of marketing content, and a post-hoc attribution method provides marketers with actionable insights to improve content in specific marketing areas. These insights help identify the strengths and weaknesses of existing content and provide evidence-based design recommendations. The study demonstrates that the insight scoring model and the insights it generates are effective, yielding promising results in both quantitative and qualitative analyses.

Agriculture companies are under significant financial pressure, and efficient decision-making is imperative. As stated by , the digital marketing analytics are significantly related to such significant agroeconomic indexes as the Agriculture Employment Rate, the Chemical Product Price Index, the Farm Product Price Index, and the Machinery Equipment Price Index. Using regression and correlation analysis, we examined data from the websites of five agricultural companies and determined the importance of various relationships. The results revealed that these indexes showed strong correlations with metrics such as branded traffic, social and search sources, and marketing costs. The study points out that to make better investment and strategic choices, agricultural companies can invest more in digital marketing analytics and artificial intelligence, which will enable them to make better decisions by better understanding trends in employment and price changes within their industry.

In predictive modeling, firms often encounter high-dimensional data encompassing multiple channels, websites, demographics, purchase types, and product categories. In the view of [11], traditional customer response models depend significantly on feature engineering, with their success hinging on the analyst's domain knowledge to develop effective predictors. As data complexity increases, these conventional models become increasingly intricate. This study demonstrates that deep learning, particularly long-short term memory neural networks, which utilize only raw data as input, can accurately predict customer behavior. In an initial application, the model surpasses standard benchmarks, while in a second, more complex scenario, it outperforms 269 out of 271 manually constructed models with diverse features and approaches, often by a substantial margin. The findings position long-short term memory neural networks as strong candidates for customer behavior modeling using panel data in complex contexts, such as direct marketing, brand choice analysis, clickstream data, and churn prediction.

Restaurateurs rely on the overall ratings of reviews on platforms like Yelp, Google, and TripAdvisor to manage the customer experience. According to [12], the main challenge lies in identifying specific aspects of the restaurant that need improvement through in-depth analysis of reviews. This study introduces a novel aspect-based deep learning framework aimed at enhancing the customer experience by extracting Key Performance Indicators (KPIs) from the sentiment of restaurant reviews. The framework integrates an information retrieval algorithm, Okapi BM25, with a deep learning model, word2vec-cnn, and is trained on a dataset of 600,000 Yelp reviews. The research identifies KPIs linked to five aspects: flavor, cost, ambience, hygiene, and service, to guide restaurateurs in improving customer experience. The findings indicate that diners generally express positive sentiment regarding "service" but negative sentiment concerning "cost." The proposed framework achieved 94% accuracy and an AUROC of 0.98, demonstrating that this innovative model can effectively convert unstructured customer feedback into actionable KPIs, supporting restaurateurs in refining their service and overall experience.

Customer sentiments are essential for business success, as they directly affect product sales, market adoption, and overall product viability. Following the work of [13], social media platforms like Facebook, with 2.32 billion monthly active users, and Twitter, with 126 million, offer a powerful means for companies to capture and analyze customer opinions. If these sentiments are not accurately assessed, it can result in product failures and significant harm to a company's reputation. Traditional methods, such as manual surveys and static report generation, are labor-intensive and lack the ability to provide real-time insights. Social media, on the other hand, facilitates the collection of candid and immediate consumer feedback. The cognitive service model is particularly effective, analyzing text and assigning sentiment scores from 0 (negative) to 1 (positive), thereby detecting sentiment across platforms, including social media, customer reviews, and discussion forums. This approach helps companies evaluate product reception during the prototype stage and make necessary modifications based on user feedback before a full-scale launch. The system is implemented using Azure services, including Logic Apps, Cognitive Services, SQL Database, and App Services, and uses Power BI for generating real-time, interactive business intelligence reports.

Payment data is one of the most valuable assets that retail banks can leverage to gain a competitive edge over new market entrants, such as Fintech companies and large internet corporations. In the research conducted by [14], the value of data in marketing is linked to its ability to reveal customer preferences: the more insight a company has into its customers, the more effective its marketing strategies become. This study introduces a business-to-business-to-consumer lead generation application built on payment transaction data within the online banking system. In this approach, the bank acts as an intermediary between private customers and merchants, employing machine learning-driven marketing strategies to create a lead generation tool that enables merchants to launch data-driven campaigns through banking channels to reach retail customers.

The introduction of machine learning into marketing practices has become vital to contemporary business. Within the analysis of the article by , scientific manuscripts that are indexed in the Scopus database are examined to learn how this integration is being carried out. The study is characterized by a targeted search for academic papers that include both machine learning and marketing in their titles, thereby yielding a comprehensive set of relevant papers. The Supabase platform was used to process these papers, with tasks including text refinement and feature extraction. Two significant machine learning approaches, topic modeling based on non-negative matrix factorization and comparative analysis using the k-means clustering algorithm, were also used in the study.

Data mining and machine learning methods can enhance the effectiveness of digital marketing strategies. In the research conducted by [15], the author highlights the capabilities of business analytics, emphasizing how data and application integration solutions transform raw data into actionable knowledge and provide more adaptable data analysis tools. The study further defines the key characteristics of Big Data Technologies, providing clarity on this topic.

Exponential technological growth provides significant opportunities for leveraging data-driven approaches in digital marketing practices. In the view of [16], machine learning can forecast future trends and facilitate decision-making by extracting valuable insights from vast data sets, greatly influencing and optimizing organizational strategic decision-making. However, a gap in research reveals limited understanding of marketers' attitudes toward, and knowledge of, machine learning tools, as well as their adoption and application for supporting strategic and operational management.

The exploration of organizational resources, competencies, and capabilities necessary for successful machine learning development projects in marketing operations reveals significant insights. In the research conducted by , the Agile-Stage-Gate model was used to examine the structure of such projects, mapping the workflow, tasks, and roles of marketing management and development teams.

Digital marketing encompasses the promotion, sale, and distribution of products or services through online platforms and channels. In the view of [17], the integration of advanced big data analytics and artificial intelligence into digital marketing is crucial for fostering sustainable marketing practices.

Recently obtained data has been used to develop an interpretive machine learning model to analyze the impact of various marketing campaigns and budgetary allocations on customer mall traffic. A dataset was collected for the analysis presented in the paper by [18] on a scale of 25 malls using AI chip-based sensors over two years to generate a large sample, which was then supplemented with campaign-specific details to enable a comprehensive analysis.

The systematic use of machine learning and data envelopment analysis is applied to analyze Twitter messages and financial metrics to identify influential social media messaging. In the study by [19], automated machine learning is employed to classify tweets from the chosen furniture retailing stores in the USA, and data envelopment analysis models are used to assess the efficiency ranking of the chosen brands using a variety of input metrics.

The constant change of the business environment is one of the most critical issues of digital marketing. The strategy that worked yesterday might lose traction over time, while the strategy that was ignored yesterday becomes relevant again, as explained in the research by . The paper presents a reinforcement learning model that leverages real-world data to enhance the performance of online marketing campaigns, overcoming the drawbacks of conventional A/B testing.

Revenue models are a critical component of strategic business decisions. As evidenced in the study by [20], the integration of machine learning with multi-case theory building sheds light on determining the optimal revenue model for a diverse range of App Store products.

Uplift modeling has become a pivotal tool in online marketing. As outlined in the research by [21], revenue uplift modeling offers greater potential by directly linking outcomes to corporate income. The study introduces a novel approach leveraging zero-inflated lognormal loss for response regression and developing a corresponding adaptable network architecture.

The concept of Internet plus agriculture was introduced in 2015 through the Chinese government's work report. As reported by , this integration has streamlined the sales process by reducing intermediaries, leading to the emergence of numerous e-commerce professional villages and a surge in rural e-commerce stores across China.

The widespread adoption of digital content has significantly altered media consumption patterns. As demonstrated by [22], the rise in online video consumption has led to a growing market for online video advertising, where the effectiveness of advertisements relies on attracting a large number of viewers. The study introduces a deep learning-based model to improve the accuracy of advertising inventory predictions by analyzing raw data from online video channels.

3. MATERIALS AND METHODS

The suggested methodological framework of stock price prediction is shown in Figure 1. The process begins by gathering the raw dataset, which is then subjected to data preprocessing procedures, such as handling missing and null values, normalization, and feature encoding, to ensure data quality and consistency. After preprocessing, the dataset is split into training (80%) and testing (20%) sets to develop and validate the model. Several baseline models are applied, including XGBoost, Support Vector Regression, Decision Tree, and Linear Regression, to provide a benchmark for comparative performance. This is followed by feature selection, which is done using advanced optimization algorithms such as Dipper Throated Optimization (DTO), Whale Optimization Algorithm (WOA), Particle Swarm Optimization (PSO), Grey Wolf Optimization (GWO), and Genetic Algorithm (GA). The DTO algorithm is also used to optimize model parameters and predictive accuracy. Lastly, the models are also assessed using relevant statistical and error-based measures to determine the best configuration.

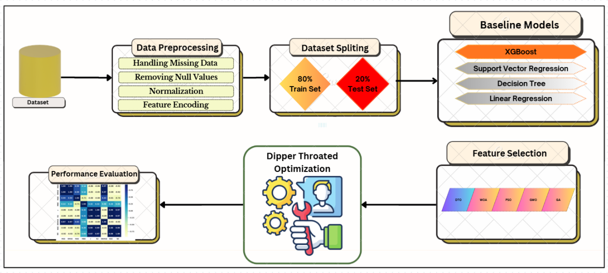

Figure 1. Proposed framework for stock price prediction incorporating data preprocessing, feature selection, baseline models, and Dipper Throated Optimization (DTO) for performance enhancement.

4. DATASET DESCRIPTION

The dataset employed in this study is the *Digital Marketing Metrics* dataset, which captures daily marketing performance indicators across multiple advertising campaigns and channels. The data provides an opportunity to analyze the relationship between marketing expenditure and resulting revenue outcomes, thereby enabling the assessment of marketing efficiency and return on investment.

The primary purpose of the dataset is to analyze marketing spending across campaigns and extract actionable insights that can guide optimization of marketing strategies. Specifically, the dataset supports the following analytical objectives:

1. **Overall Return on Marketing Investment (ROMI):** Evaluate the overall effectiveness of marketing expenditures in generating revenue.
2. **ROMI by Campaigns:** Assess the performance of individual campaigns to identify which efforts yield the highest profitability.
3. **Temporal Performance Analysis:** Examine campaign performance across dates, identifying periods of highest spending, peak revenue generation, and variations in conversion rates and average order values.
4. **Customer Activity by Day:** Compare average revenue across weekdays and weekends to determine when buyers are most active.
5. **Campaign Type Effectiveness:** Identify which campaign types—such as social media, banner ads, influencer marketing, or search engine campaigns—achieve superior results.
6. **Geographical Performance:** Evaluate performance differences across tier 1 and tier 2 cities to determine optimal targeting regions.

The dataset consists of several variables that describe campaign-level marketing performance on specific dates. Each row in the dataset corresponds to a single marketing campaign's activity for a given day. The variables are defined as follows:

Table 1. Description of Dataset Attributes

Column	Description
`featureid`	Unique identifier for each record in the dataset.
`c_date`	Date on which the marketing expenditure occurred.
`campaign_name`	Name or description of the marketing campaign.
`category`	Type or channel of the campaign (e.g., social, search, influencer, banner).
`campaign_id`	Unique identifier assigned to each campaign.
`impressions`	Number of times an advertisement was displayed to users.
`mark_spent`	Amount of money spent on the campaign for that specific day.
`clicks`	Number of times users clicked on the ad or visited the website.
`leads`	Number of users who provided contact information or registered interest.
`orders`	Number of successful transactions or purchases made.
`revenue`	Total revenue generated by the campaign on that day.

The dataset records marketing activities on a daily basis across multiple channels. For each marketing campaign on a specific date, the data captures the full marketing funnel—from impressions and clicks to leads, orders, and generated revenue. For example, for the campaign *facebook_tier1* on February 1st, the company spent INR 7,307.37 and recorded 148,263 impressions. These impressions resulted in 1,210 clicks, which in turn led to 13 leads and 1 finalized order, generating a total revenue of INR 4,981.

This data provides a factual account of marketing activity, answering questions such as: how much was spent, how much was earned, and how customers interacted with the advertisements. The inclusion of variables such as impressions, clicks, leads, and revenue enables the computation of essential marketing performance indicators, including:

- **Click-Through Rate (CTR)** – measures user engagement through clicks per impression.

- **Conversion Rate (CR)** – quantifies how effectively clicks convert into leads or orders.
- **Cost per Click (CPC)** – indicates spending efficiency relative to user engagement.
- **Return on Marketing Investment (ROMI)** – evaluates profitability relative to marketing spend.
- **Average Order Value (AOV)** – measures the mean transaction value for each order.

These derived metrics are crucial for assessing campaign performance and comparing marketing strategies across different channels and time periods. By leveraging these indicators, the study aims to build a predictive framework that identifies the factors most strongly correlated with revenue generation. The dataset allows for the exploration of multiple dimensions of marketing performance, including:

1. **Temporal Trends:** Understanding how spending patterns and conversion outcomes vary over time.
2. **Channel Efficiency:** Determining which advertising media deliver the highest revenue relative to cost.
3. **Geographic Segmentation:** Evaluating campaign effectiveness across different city tiers or market regions.
4. **Behavioral Insights:** Identifying when customers are most active and how user behavior influences conversion outcomes.

In general, this data is an in-depth prerequisite for implementing sophisticated machine learning and optimization algorithms, including the Dipper Throated Optimization (DTO)-improved XGBoost model, which would predict revenue and enhance the marketing decision-making process. A correlation analysis was also used to examine relationships among key performance metrics, including impressions and clicks, revenue, cost per click (CPC), customer acquisition cost (CAC), and return on marketing investment (ROMI). As illustrated in Figure 2, the correlation heatmap provides a graphical overview of these interactions, including both positive and negative relationships. The positive correlations, e.g., between clicks and revenue, are strong, indicating that higher user engagement is associated with higher revenue. On the other hand, the negative coefficient between CAC and ROMI indicates that higher acquisition costs are likely to lower overall marketing efficiency.

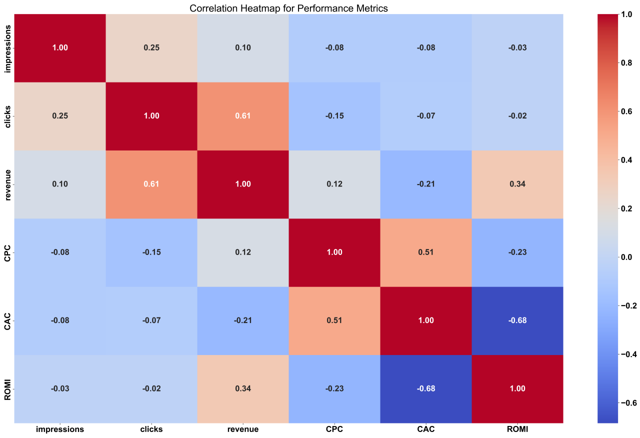

Figure 2. Correlation Heatmap for Performance Metrics

5. DATA PREPROCESSING

Data preprocessing represents a crucial stage in the data analysis pipeline, particularly when preparing raw marketing data for machine learning applications. The accuracy and reliability of any predictive model are heavily dependent on the quality and consistency of the data it learns from. Since real-world marketing datasets often contain inconsistencies, missing values, outliers, and variables with differing scales, rigorous preprocessing ensures that the resulting models are both statistically sound and computationally efficient. Real-world data are seldom perfect. Missing, incomplete, or invalid entries may arise due to human error, system logging failures, or incomplete customer responses. Such irregularities can distort statistical summaries and degrade model performance if left untreated. Therefore, the dataset was carefully examined to detect and handle missing values across key features such as `impressions`, `clicks`, `leads`, `orders`, and `revenue`. Two main strategies were employed:

1. **Deletion of Irrelevant or Highly Incomplete Records:** Observations containing excessive missing or erroneous information (e.g., more than 30% missing fields) were removed. This approach minimizes noise and prevents unreliable imputations from affecting model training.
2. **Imputation of Missing Values:** For numerical features with minor gaps, statistical imputation techniques were applied. Continuous variables such as `mark_spent`, `impressions`, and `revenue` were imputed using the median of their respective distributions, ensuring robustness against outliers. For categorical features such as `category` or `campaign_name`, the most frequent category (mode) was substituted.

Additionally, all feature columns were validated for logical consistency. For instance, `clicks` could not exceed `impressions`, and `orders` could not exceed `leads`. Any records violating these logical constraints were corrected or removed. Outliers in numerical columns—often resulting from data entry or measurement errors—were identified using the Interquartile Range (IQR) method and treated appropriately. This ensured that only valid, meaningful values were retained for analysis. The scale of features is also essential when feeding machine learning algorithms like XGBoost, because different variables may have significant numerical differences. An example is of having impressions in the hundreds of thousands but clicking and orders in the tens or hundreds. In the absence of scaling, attributes with large values can dominate during model training and lead to biased predictions. Normalization puts feature values on a standard scale without distorting their range. This research utilized the MinMax Normalization of all continuous numerical variables to bring them under a range of [0,1], as expressed in Equation (1):

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

where:

- x is the original feature value,
- $x_{\min}$ and $x_{\max}$ represent the minimum and maximum observed values of that feature, and
- x' is the normalized output value.

Such a transformation ensures that every input feature is applied proportionately to the learning process, increasing the convergence rate and stability of the model during optimization. Moreover, this is because normalization benefits gradient-based algorithms, such as XGBoost, which are theoretically and numerically stable against gradient explosions from high-magnitude features. In the exploratory data analysis, the distributions of the key variables (mark spent, clicks, leads, orders, and revenue) were plotted before and after normalization. The comparison demonstrated the transformation's ability to scale feature ranges whilst maintaining the general distributional properties. Encoding was also performed to prepare categorical variables, such as category and campaign name, as model inputs: ordinal variables were encoded with label encoding, and non-ordinal categories were encoded with one-hot encoding, which converts each campaign type into a binary representation. This transformation ensures that categorical variables are expressed numerically without imposing any artificial ranking. Lastly, all the features were normalized into a single dataset, which was then trained and optimized to produce a model. To ensure unbiased evaluation of the models, the processed data were split into training and test sets at 80:20. This purified and normalized dataset will serve as the basis for further modeling using the XGBoost framework optimized with the DTO.

6. MACHINE LEARNING MODELS

Machine learning methods are emerging as effective for addressing complex forecasting challenges, especially in environments with large amounts of data, such as digital marketing. The choice of suitable models is an important consideration, as both the time dynamics and the nonlinear relationships between variables are reflected in the predictive framework. The models selected for this research are XGBoost, Support Vector Regression (SVR), Decision Tree (DT), and Linear Regression (LR), each offering distinct benefits for revenue prediction. The selection of models was based on two main criteria, namely (1) their applicability to the sequence and time-dependent modeling, and (2) their successful experience in the digital marketing and financial forecasting tasks. Nonlinear relationships, seasonal changes, and lagged relationships between marketing spending and earned revenue are typically observed in temporal marketing data. Thus, the models used should be able to describe both linear trends and complex nonlinear interactions. Moreover, these models remain popular in the marketing analytics and econometrics literature because they are easily interpretable, scalable, and capable of handling noisy, high-dimensional data. In particular:

- **Relevance to Temporal Modeling:** Models were required to handle sequential observations over time, allowing for the recognition of patterns such as campaign seasonality or daily fluctuations in consumer engagement.
- **Predictive Accuracy and Efficiency:** Algorithms were evaluated for their ability to balance prediction precision with computational feasibility when applied to large-scale marketing datasets.
- **Popularity and Proven Use in Literature:** All selected models are well-documented in prior studies on digital marketing optimization, customer behavior prediction, and revenue forecasting.

6.0.1 Extreme Gradient Boosting (XGBoost)

XGBoost is an ensemble learning algorithm based on the gradient boosting framework, designed to achieve high predictive performance and computational efficiency. It builds a series of decision trees, with each subsequent tree improving on its predecessor by minimizing a differentiable loss function. To avoid overfitting, regularization parameters are added, and to scale to large datasets, parallelized tree construction is used. The model prediction is mathematically as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathscr{F} \tag{2}$$

And where the regression trees space is denoted by the letter $\mathscr{F}$, and K is the number of trees. The fact that XGBoost can model nonlinear and hierarchical interactions makes it especially appropriate for temporal marketing data, where feature relationships change dynamically over time. Also, since it is compatible with optimization algorithms such as Dipper Throated Optimization (DTO), hyperparameter tuning can be automated, enhancing overall performance.

6.0.2 Support Vector Regression (SVR)

Support Vector Regression is an extension of Support Vector Machines (SVMs) to regression problems. It works by locating a function that best approximates the relationship between input features and target values while maintaining an acceptable level of error, denoted by ε, relative to the regression hyperplane. The decision role of the model is presented as follows:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{3}$$

where $K(x_i, x)$ is the kernel function that maps inputs into a high-dimensional feature space, and α_i, α_i^* are Lagrange multipliers. SVR is particularly effective for small to medium-sized datasets with nonlinear dependencies, as it can utilize kernel functions (e.g., radial basis function, polynomial) to capture complex relationships between marketing inputs (such as impressions or spending) and outputs (such as revenue). However, SVR can be computationally intensive for very large datasets and requires careful kernel and hyperparameter tuning.

6.0.3 Decision Tree Regression (DT)

A Decision Tree is a non-parametric supervised learning algorithm that divides the feature space into discrete clusters around a threshold of the features. At each node, the algorithm selects the feature and threshold that maximize information gain or minimize prediction error (e.g., mean squared error). The model recursively splits the data until terminal nodes (leaves) represent homogeneous subsets. The prediction for a given observation is computed as the mean value of target variables in the corresponding leaf node. Decision Trees offer interpretability and flexibility, making them suitable for identifying key drivers of marketing performance. However, they may overfit noisy data unless pruned or regularized. When used as base learners in ensemble models such as XGBoost, they contribute significantly to predictive accuracy and robustness in temporal modeling.

6.0.4 Linear Regression (LR)

Linear Regression serves as a baseline model to benchmark the performance of more sophisticated algorithms. It assumes a linear relationship between the dependent variable (revenue) and independent variables (e.g., marketing budget, impressions, clicks), expressed as:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i \tag{4}$$

where β_0 is the intercept, β_j are the regression coefficients, and ε_i represents the residual error. Despite its simplicity, Linear Regression remains a useful interpretive tool for quantifying direct relationships and estimating marginal effects. However, its assumption of linearity limits its capacity to model complex temporal and nonlinear dependencies often present in digital marketing data. Among the chosen algorithms, XGBoost is the most appropriate for temporal models because it can handle sequential dependencies, nonlinear feature interactions, and variable significance that varies over time. Its boosting logic can be used to provide iterative error correction and flexibility to changing marketing behaviors. Decision Trees offer interpretability and transparency, whereas Support Vector Regression provides robustness to smooth, nonlinear patterns. Linear Regression, whilst less flexible, provides an initial platform for understanding linear trends and variable impact, serving as a point of comparison with more sophisticated models. In combination, these models allow exploring both linear and nonlinear temporal relationships in the marketing data in a one-on-one manner, enabling the determination of the most effective modeling framework for accurately forecasting revenue.

6.1 Proposed Dipper Throated Optimization (DTO)

The Cinclidae birds are known to bob or dip when they are sitting in a particular place, like the Dipper Throated bird. The characteristic that makes this bird unique in comparison to other passerines is the fact that it can dive, swim and hunt beneath the surface of the water. Its short and flexible wings allow it to fly straight and fast without having to stop and glide. The hunting behavior is characterized by very rapid bowing, and the white breast is unique to the Dipper-Throated bird. It dives wildly among the waves of a small waterfall, or the rush in of a large stream, where pebbles and stones are seized by it as it darts about, and the little fish and invertebrates that have their homes below the surface are startled by the splash of the water and blown out of their hiding-places. This foraging is naturally described by the Dipper Throated Optimization (DTO) algorithm, which models the behavior of a flock of birds ducking and diving in search of food. A bird will be sent to explore the global optimum, with its position and velocity iteratively updated to explore possible solutions. The algorithm is based on simulating both the exploratory diving (search) and the exploitative flight (refinement). The positions and velocities of the birds in the population can be expressed as matrices, as shown below:

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} & \dots & A_{1,d} \\ A_{2,1} & A_{2,2} & A_{2,3} & \dots & A_{2,d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & A_{m,3} & \dots & A_{m,d} \end{bmatrix} \tag{5}$$

$$B = \begin{bmatrix} B_{1,1} & B_{1,2} & B_{1,3} & \dots & B_{1,d} \\ B_{2,1} & B_{2,2} & B_{2,3} & \dots & B_{2,d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B_{m,1} & B_{m,2} & B_{m,3} & \dots & B_{m,d} \end{bmatrix} \tag{6}$$

Here, $A_{(i,j)}$ denotes the position of the i^{th} bird in the j^{th} dimension for $i \in \{1,2,\dots,m\}$ and $j \in \{1,2,\dots,d\}$. Similarly, $B_{(i,j)}$ represents the bird's velocity in the j^{th} dimension. The initial positions $A_{(i,j)}$ are uniformly distributed within the search space.

The fitness values of each bird are computed as:

$$h = \left[h_1(A_{1,1}, A_{1,2}, \dots, A_{1,d}), h_2(A_{2,1}, A_{2,2}, \dots, A_{2,d}), \dots, h_m(A_{m,1}, A_{m,2}, \dots, A_{m,d}\right. \tag{7}$$

Each bird's fitness reflects its success in finding food, where a higher value represents a better solution. The bird with the best fitness is designated as the *mother bird* or leader, A_{best}, while the others follow as normal birds A_{nd}. The globally best solution found so far is represented by A_{Gbest}.

The position and velocity updates are performed according to the following equations:

$$A(i+1) = \begin{cases} X, & \text{if } R < 0.5 \\ Y, & \text{otherwise} \end{cases} \tag{8}$$

$$X = A_{\text{best}}(i) - K_1 \left| K_2 \cdot A_{\text{best}}(i) - A(i) \right| \tag{9}$$

$$Y = A(i) + B(i+1) \tag{10}$$

$$B(i+1) = K_3 B(i) + K_4 r_1 (A_{\text{best}}(i) - A(i)) + K_5 r_2 (A_{Gbest} - A(i)) \tag{11}$$

where:

- i is the iteration number,
- $A(i)$ is the position of the bird at iteration i,
- $A_{\text{best}}(i)$ is the best local solution at iteration i,
- A_{Gbest} is the global best solution found so far,
- $B(i+1)$ is the velocity at iteration $i+1$,

- K_1, K_2, K_3 are weight coefficients controlling convergence,
- K_4, K_5 are constants influencing attraction to local and global bests,
- r_1, r_2 are random values in the range $[0,1]$,
- R is a random threshold determining exploration vs. exploitation.

The algorithm alternates between diving (exploration) and flying (exploitation) behaviors depending on the random parameter R. When $R < 0.5$, the birds dive deeper (exploration phase), while $R \geq 0.5$ represents flying movements to refine their positions (exploitation phase).

Algorithm 1 Dipper Throated Optimization (DTO) Algorithm

Initialize bird positions BP_i, $i = 1, 2, \ldots, n$
Initialize velocities BV_i, total iterations T_{max}
Define fitness function f_n, parameters $c, C_1, C_2, C_3, C_4, C_5, r_1, r_2, R$, and set $t = 1$
Evaluate f_n for each bird BP_i
Find best bird BP_{best}
while $t \leq T_{max}$ **do**
  for $i = 1$ **to** n **do**
    if $R < 0.5$ **then**
      Update position of swimming bird:
      $BP_{nd}(t+1) = BP_{best}(t) - C_1 |C_2 BP_{best}(t) - BP_{nd}(t)|$
    else
      Update velocity:
      $BV(t+1) = C_3 BV(t) + C_4 r_1 (BP_{best}(t) - BP_{nd}(t)) + C_5 r_2 (BP_{Gbest} - BP_{nd}(t))$
      Update position of flying bird:
      $BP_{nd}(t+1) = BP_{nd}(t) + BV(t+1)$
    end if
  end for
  Recalculate f_n for all birds
  Update parameters c, C_1, C_2, R
  Find new BP_{best} and set $BP_{Gbest} = BP_{best}$
  $t = t + 1$
end while
Return global best bird BP_{Gbest}

The DTO algorithm thus simulates the dual behavior of the Dipper Throated bird—rapid underwater dives for exploration and swift aerial maneuvers for exploitation. Its ability to balance these phases enables it to converge efficiently toward optimal solutions in complex search spaces, making it suitable for applications such as feature selection, parameter tuning, and predictive modeling.

7. EMPIRICAL RESULTS

The results of the experiment in which the selected machine learning models were applied to the processed digital marketing data are presented in this section. The purpose of this analysis is to compare the predictive performances of the two models in marketing revenue prediction based on campaign characteristics and campaign spending behavior. A series of experiments designed to compare baseline models in a variety of statistical and error-based performance indicators, including the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), the Mean Bias Error (MBE), the coefficient of correlation (r), the coefficient of determination (R^2), Relative Root Mean Square Error (RRMSE), Nash Sutcliffe Efficiency (NSE), and the Index of Willmott (WI). All these measurements are used to determine the accuracy and consistency.

7.1 Baseline Model Comparison

Table 2 shows the pure performance indicators of the four machine learning models, including XGBoost, Support Vector Regression (SVR), Decision Tree Regression (DT) and Linear Regression (LR). The outputs of both models were tested on the test data, providing insight into how they generalized beyond the training data used.

Table 2. Baseline Performance Metrics for All Models

Model	MSE	RMSE	MAE	MBE	r	R^2	RRMSE	NSE	WI
XGBoost	0.0905	0.3008	0.2555	0.0236	0.7881	0.8007	21.6542	0.8298	0.8345
Support Vector Regression	0.1086	0.3610	0.3066	0.0283	0.6305	0.6431	22.6144	0.8071	0.6676
Decision Tree	0.1267	0.4212	0.3577	0.0330	0.4729	0.4855	22.4262	0.6980	0.5007
Linear Regression	0.1448	0.4813	0.4088	0.0377	0.3153	0.3279	25.6421	0.5018	0.3338

The comparative analysis shows that the XGBoost model is more effective than all other models across most performance indicators. It had the lowest MSE (0.0905), RMSE (0.3008), and MAE (0.2555), indicating the lowest average prediction error. Besides, XGBoost achieved the best R^2 (0.8007) and correlation coefficient (r = 0.7881), indicating a strong linear relationship between the predicted and actual revenue values. Support Vector Regression (SVR) was also reasonably good, with moderate low error and reasonable correlation values, confirming that it can capture nonlinear relationships. Nevertheless, it is less efficient than XGBoost on large datasets because the former is less sensitive to computational cost and to kernel parameter settings. The Decision Tree and Linear Regression models were less effective at capturing the nonlinear structure of marketing data, resulting in higher errors and lower correlation scores. The findings justify using XGBoost as the main model, which will be further fine-tuned and optimized. It is also appropriate to forecast in changing marketing scenarios due to its integration of ensemble learning, regularization and scalability. The following section, therefore, aims to enhance its predictive accuracy by using the Dipper Throated Optimization (DTO) algorithm to optimize hyperparameters. To compare the overall consistency and variability in the performance of multiple predictive models, a parallel coordinates visualization was used. According to Figure 3 this plot shows the normalized values of some evaluation measures, such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), correlation coefficient (r), coefficient of determination (R^2), Relative RMSE (RRMSE), Nash Sutcliffe Efficiency (NSE), and Willmott Index (WI). Including the mean and standard deviation lines provides a statistical reference point for the relative stability of each model. The XGBoost model offers the best balance between accuracy and consistency, as its metric values are close to or below the mean across most dimensions. Hierarchical clustering was also used to explore further the relationships

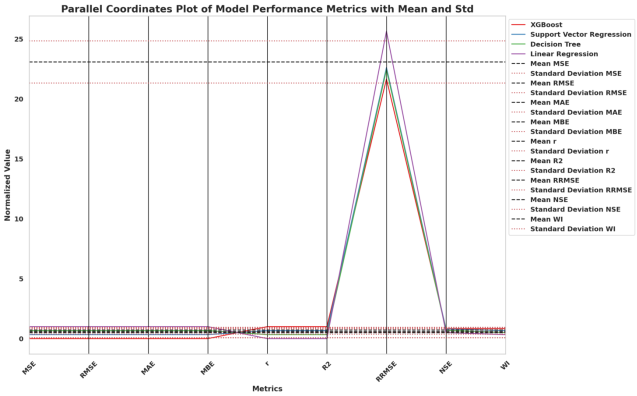

Figure 3. Parallel Coordinates Plot of Model Performance Metrics with Mean and Standard Deviation

between the predictive models and their performance similarities. The dendrogram shows how other models were grouped according to the distance measure and the metric correlations as seen in Figure 4. The clustering indicates that both Decision Tree and Linear Regression exhibit similar performance traits and form a cluster, while XGBoost and Support Vector Regression are closely related in another cluster. This hierarchical structure provides a valuable understanding of model performance, suggesting that ensemble and kernel approaches exhibit performance characteristics different from those of simple tree-or regression-based approaches.

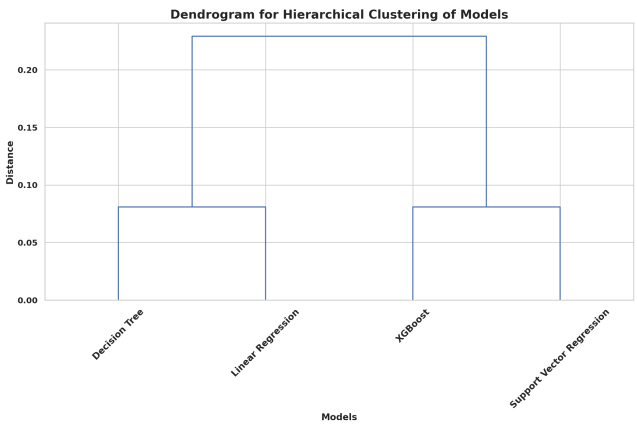

Figure 4. Dendrogram for Hierarchical Clustering of Models

A radar chart was used to provide a multidimensional view of model performance across the main evaluation metrics. This chart, as shown in Figure 5, is a concurrent representation of many values- Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), correlation coefficient (r), coefficient of determination (

$$R^2$$

) and Relative RMSE (RRMSE) to each predictive model. The radar plot shows that although all the models exhibit similar general trends, XGBoost and Support Vector Regression tend to be slightly more effective at reducing errors and finding the optimal balance between correlation, compared to Decision Tree and Linear Regression, which supports the interdependence of their relative strengths and precision.

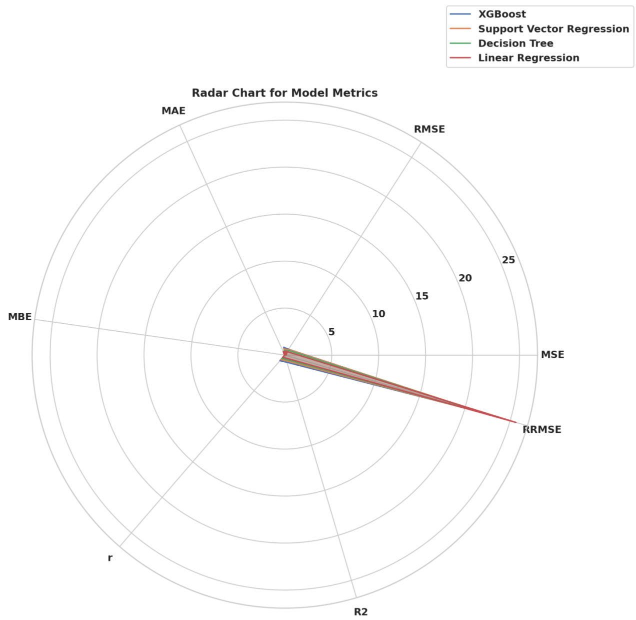

Figure 5. Radar Chart for Model Metrics

7.2 XGBoost Optimization Results

Once the baseline performance of the XGBoost was identified, several metaheuristic algorithms were applied to improve the model's predictive capacity further. The task of each of the optimizers was to optimize several important hyperparameters, including the learning rate, the maximum tree depth, the subsample ratio, and the number of estimators, to the extent that they would reduce the amount of prediction error and achieve as high a correlation between the predicted and observed values of revenue as possible. The optimization methods applied included the Dipper Throated Optimization (DTO) algorithm, the Whale Optimization Algorithm (WOA), the Particle Swarm Optimization (PSO), the Grey Wolf Optimizer (GWO), and the Genetic Algorithm (GA). The purpose of comparing these optimization strategies was to determine which approach yields the most stable and accurate model for digital marketing revenue forecasting. Table 3 reports the performance metrics obtained for each optimized XGBoost variant. Metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), correlation coefficient (r), coefficient of determination (R^2), Relative Root Mean Square Error (RRMSE), Nash–Sutcliffe Efficiency (NSE), and Willmott's Index (WI) were used to assess both accuracy and consistency.

Table 3. Performance Metrics for XGBoost Optimization using Different Metaheuristic Algorithms

Model	MSE	RMSE	MAE	MBE	r	R^2	RRMSE	NSE	WI
XGBoost + DTO	0.0010	0.0034	0.0029	0.0016	0.8970	0.9002	8.6498	0.9298	0.9736
XGBoost + WOA	0.0123	0.0410	0.0348	0.0032	0.8576	0.8608	11.1797	0.8571	0.7789
XGBoost + PSO	0.0144	0.0479	0.0406	0.0037	0.8182	0.8214	11.1016	0.7979	0.5842
XGBoost + GWO	0.0165	0.0547	0.0465	0.0043	0.7788	0.7820	12.6430	0.6818	0.3894
XGBoost + GA	0.0195	0.0645	0.0565	0.0005	0.6788	0.6820	13.1453	0.5722	0.2107

The results in Table 3 clearly show that the **DTO-optimized XGBoost model** outperforms all other metaheuristic-enhanced models across nearly all performance metrics. It achieved the lowest error values—MSE (0.0010), RMSE (0.0034), and MAE (0.0029)—indicating superior accuracy and minimal deviation from observed values. Additionally, DTO yielded the highest correlation coefficient ($r = 0.8970$) and coefficient of determination ($R^2 = 0.9002$), confirming a powerful linear relationship between predicted and actual revenue. Furthermore, the DTO model achieved the highest Nash–Sutcliffe Efficiency (0.9298) and Willmott's Index (0.9736), both indicators of outstanding model reliability and predictive consistency. These metrics validate that the model optimized through DTO not only captures the underlying nonlinear dependencies in the data but also generalizes effectively to unseen cases. In comparison, the WOA and PSO optimizers achieved reasonably good performance, reducing prediction errors relative to the unoptimized baseline but still falling short of DTO's accuracy and correlation strength. The GWO and GA approaches performed the worst, with higher error rates and lower correlation measures, suggesting limited exploration of the parameter space and premature convergence. In general, these results show that the Dipper Throated Optimization (DTO) algorithm is more effective and stable for hyperparameter optimization of the XGBoost model. Its adaptive exploration-exploitation trade-off enables it to explore near-optimal configurations, resulting in considerable reductions in prediction errors and considerable increases in model robustness. As such, the DTO-optimized XGBoost model is confirmed to be the most effective and precise forecasting model for predicting revenue in digital marketing applications. A box-and-whisker plot was used to visualize the variability and dispersion of the model performance measures for individual models. Figure 6 shows a detailed chart of the distribution patterns of popular evaluation metrics across models, such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), correlation coefficient (r), coefficient of determination (

$$R^2$$

), Relative RMSE(RRMSE), NashSutcliffe Efficiency (NSE) and Willmott Index (WI). The box plots reflect the central tendency and dispersion of each metric, and the overlaid swarm plots show the individual data points, highlighting outliers and the consistency of model performance. It is a visualization that clearly shows model stability and comparative correctness.

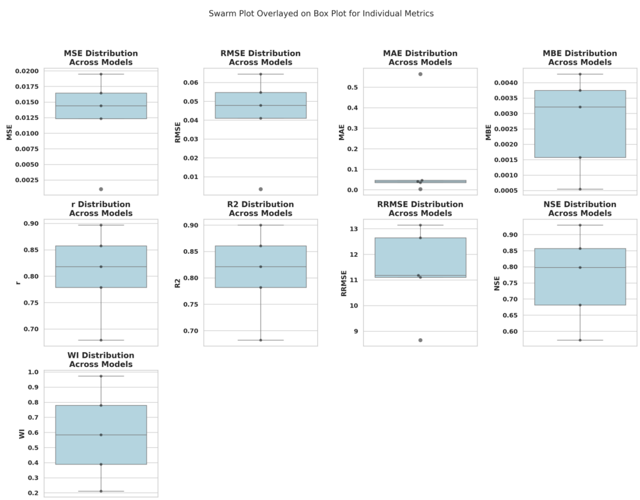

Figure 6. Swarm Plot Overlayed on Box Plot for Individual Metrics

A visual study of the most critical metrics—errors and efficiency — was used to evaluate the relative performance of XGBoost models optimized with various metaheuristic algorithms. The comparison involves XGBoost combined with Differential Tuned Optimization (DTO), the Whale Optimization Algorithm (WOA), Particle Swarm Optimization (PSO), and the Genetic Algorithm (GA), as shown in Figure 7. The variations in error indicators are summarized in the chart using the correlation coefficient (r), the coefficient of determination (R^2), the Relative Root Mean Squared Error (RRMSE), the Nash-Sutcliffe efficiency (NSE), and the Index of Willmott (WI). There are clear indications that the XGBoost+DTO model is the best-performing, with the lowest error rates and the highest stability across all models tested.

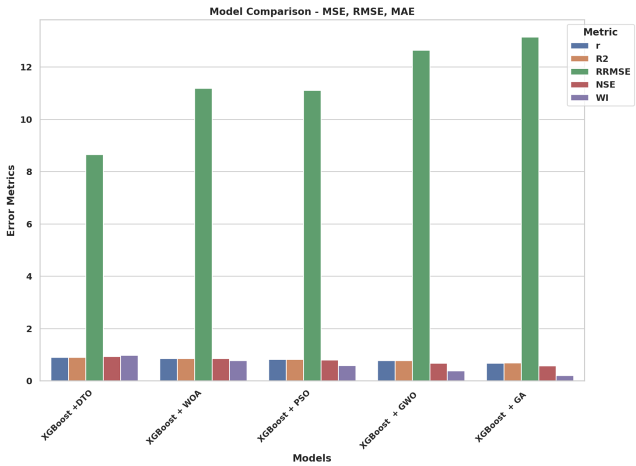

Figure 7. Model Comparison – MSE, RMSE, MAE

To summarize the central tendency and variability of the model evaluation metrics, a descriptive statistical analysis was conducted. As presented in Figure 8, the mean and standard deviation values are visualized for each performance metric, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), correlation coefficient (r), coefficient of determination (R^2), Relative RMSE (RRMSE), Nash–Sutcliffe Efficiency (NSE), and Willmott's Index (WI). The chart highlights that RRMSE exhibits the highest mean and standard deviation, indicating greater variability across models, whereas other metrics such as MSE and RMSE show relatively consistent performance. This provides a foundational overview of the overall model stability and comparative dispersion across evaluation criteria.

8. DISCUSSION

The empirical evidence demonstrates a clear hierarchy of predictive performance across modeling choices and optimization strategies. At baseline, XGBoost already outperformed Support Vector Regression, Decision Tree, and Linear Regression across nearly every metric, reflecting the advantage of boosted tree ensembles in capturing nonlinearity, interactions, and the distributional heterogeneity typical of digital marketing data. Following the metaheuristic tuning, the XGBoost optimized by DTO has significantly lower errors (e.g., MSE, RMSE, MAE) and larger agreement statistics (e.g., r, R^2, NSE, WI), suggesting both point performance and accurate representation of both variance and temporal structure.

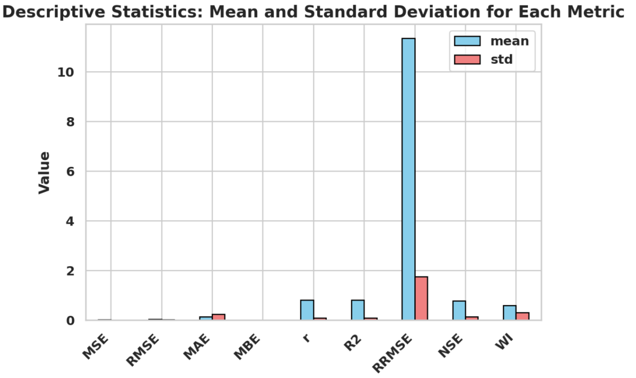

Figure 8. Descriptive Statistics: Mean and Standard Deviation for Each Metric

The level of improvement between baseline XGBoost and XGBoost+DTO indicates that search quality within the hyperparameter space, not just the model space, plays a leading role in revenue forecasting. In comparison, other optimizers (WOA, PSO, GWO, GA) showed more minor but more continuous improvements, and their error/correlation statistics were also more scattered, which is in line with their less stable exploration-exploitation characteristics. Complementary visual diagnostics (correlation heatmaps, parallel coordinates, dendrograms, radar plots, and distributional box-and-whisker summaries) all told the same tale: the DTO setup reduced error dispersion and generated narrower, more reliable prediction bands. The managerial and analytical implications of these findings are a direct contribution to budgetary allocation, campaign design, and the pace of execution. Because the feature space encodes the entire funnel (impressions $\rightarrow$ clicks $\rightarrow$ leads $\rightarrow$ orders and revenue) in addition to spend and campaign typology, the high-fidelity forecasts of XGBoost+DTO can be used to make prospective what-if-scenarios on ROMI, CPC, CAC and AOV under alternative spend schedules, channel mixes (social, banner, influencer, search) and geo-targeting strategies (tier 1 vs. tier 2). The tree structure of the model provides post-hoc interpretability (e.g., feature importance, partial dependence, SHAP-style diagnostics). It allows practitioners to assign revenue lifts to individual levers (e.g., marginal spend on high-conversion days, creative fatigue cycles, geo-channel interactions) rather than viewing the system as a black box. Methodological advice is also provided by the clustering of models (DT, LR vs. XGBoost SVR): the dynamic, nonlinear, and multi-channel data-generating processes are likely to generalize better with ensemble and kernel methods; simpler baselines are still helpful as a sanity check and as a way of communicating linear effects to the stakeholders. Combined, the uniformity and excellence of DTO-tuned XGBoost make it a viable basis for a real-time forecasting service that assists in weekly campaign planning and prompt reaction to market changes. Nevertheless, several limitations and avenues for extension warrant attention. First, generalizability depends on data scope and curation: shifts in pricing, seasonality, competitive intensity, or attribution logic can induce concept drift, so temporal cross-validation, rolling retraining, and drift monitors should be standard operating procedure. Second, while the study optimizes predictive accuracy, causality is not guaranteed; future work could integrate uplift modeling, instrumental variables, or geo–time randomized experiments to estimate incremental effects of spend on revenue. Third, leakage and target-alignment risks should be systematically audited (e.g., by ensuring that post-outcome signals do not enter the training features and by aligning normalization windows to past-only information). Fourth, external covariates (macroeconomic factors, holidays, platform algorithm changes) could further stabilize forecasts; multi-horizon variants and probabilistic calibration (pinball loss, CRPS) would add decision-ready uncertainty quantification. Finally, deployment concerns—latency, inference costs, governance, and privacy—should be baked into an MLOps pipeline, with human-in-the-loop review for significant budget moves. Addressing these considerations will strengthen robustness and help translate superior in-sample and out-of-sample accuracy into sustained financial impact.

9. CONCLUSION AND FUTURE WORK

This study presented a comprehensive data-driven framework for forecasting digital marketing revenue by integrating advanced preprocessing, comparative machine learning analysis, and metaheuristic optimization. Their results showed that XGBoost consistently outperformed Support Vector Regression, Decision Tree, and Linear Regression models by a significant margin, as it is capable of identifying nonlinear trends in marketing data. The Dipper Throated Optimization (DTO) algorithm with XGBoost at its optimal parameters reported the lowest error and the highest correlation coefficients, which proves its ability to search a larger portion of the hyperparameter space. XGBoost and DTO achieved outstanding predictive stability and accuracy, making them the most effective models for predicting marketing revenue and helping make data-driven decisions in campaign control. Visual analyses, including heatmaps, radar charts, and dendrograms, were used to support this, and the results showed consistent clustering and converging alignment of metrics, proving the DTO-optimized model. In reality, this hybrid structure enables marketers to evaluate campaign performance, conduct what-if analyses, and allocate resources with greater confidence. It helps with strategic decision-making by determining the most effective types of campaigns to run, the appropriate level of spending, and the best time to advertise. Also, the interpretability of tree-based models allows practitioners to gain insights into feature significance, relating key marketing variables — e.g., impressions, clicks, and conversion rates — to revenue performance. Altogether, the suggested solution not only improves the precision of forecasting results but also enhances the efficiency of operational processes and the strategic flexibility of digital marketing systems. Future work should expand this research in several directions. First, there is a need to integrate causal inference methods such as uplift modeling and randomized controlled experiments to move beyond prediction and toward understanding the actual incremental impact of marketing interventions. Second, incorporating temporal cross-validation, drift detection, and rolling retraining mechanisms would enhance model generalization and robustness in dynamic market conditions. Third, enriching the feature set with external factors—such as holidays, macroeconomic indicators, competitive intensity, and social trends—could further improve forecast precision. Fourth, uncertainty quantification through probabilistic and interval forecasting would allow decision-makers to manage risk and optimize budgets under uncertainty. Fifth, extending DTO into a multi-objective optimization framework could simultaneously optimize accuracy, computational cost, and interpretability. Sixth, embedding the model within an MLOps pipeline would ensure scalability, continuous integration, and compliance with governance requirements for real-time marketing analytics. Lastly, future research should conduct user-centric evaluations to assess how interpretability tools, such as SHAP and partial dependence plots, improve trust, adoption, and decision quality among marketing professionals. By addressing these aspects, the proposed framework can evolve into a fully operational, intelligent marketing decision-support system that not only predicts but also prescribes optimal actions for maximizing digital marketing performance.

DATA AVAILABILITY STATEMENT

The smart home energy consumption and weather data used in this study are publicly available at: `https://www.kaggle.com/datasets/sinderpreet/analyze-the-marketing-spending/data`.

REFERENCES

[1] S. P. Goldman, H. van Herk, T. Verhagen, and J. W. Weltevreden, "Strategic orientations and digital marketing tactics in cross-border e-commerce: Comparing developed and emerging markets," *International Small Business Journal*, vol. 39, no. 4, pp. 350–371, 2021.

[2] M. Madanchian, "Generative AI for consumer behavior prediction: Techniques and applications," *Sustainability*, vol. 16, no. 22, 2024.

[3] J. Perumalsamy, C. Althati, and L. Shanmugam, "Advanced AI and machine learning techniques for predictive analytics in annuity products: Enhancing risk assessment and pricing accuracy," *Journal of Artificial Intelligence Research*, vol. 2, no. 2, 2022, article 2.

[4] F. M. Talaat, A. Aljadani, B. Alharthi, M. A. Farsi, M. Badawy, and M. Elhosseini, "A mathematical model for customer segmentation leveraging deep learning, explainable AI, and RFM analysis in targeted marketing," *Mathematics*, vol. 11, no. 18, 2023.

[5] C. Ziakis and M. Vlachopoulou, "Artificial intelligence in digital marketing: Insights from a comprehensive review," *Information*, vol. 14, no. 12, 2023.

[6] S. Eloranta and M. Boman, "Predictive models for clinical decision making: Deep dives in practical machine learning," *Journal of Internal Medicine*, vol. 292, no. 2, pp. 278–295, 2022.

[7] A. H. Alharbi, S. K. Towfek, A. A. Abdelhamid, A. Ibrahim, M. M. Eid, D. S. Khafaga, N. Khodadadi, L. Abualigah, and M. Saber, "Diagnosis of monkeypox disease using transfer learning and binary advanced dipper throated optimization algorithm," *Biomimetics*, vol. 8, no. 3, 2023.

[8] J. R. Saura, "Using data sciences in digital marketing: Framework, methods, and performance metrics," *Journal of Innovation & Knowledge*, vol. 6, no. 2, pp. 92–102, 2021.

[9] F. Villarroel Ordenes and R. Silipo, "Machine learning for marketing on the KNIME hub: The development of a live repository for marketing applications," *Journal of Business Research*, vol. 137, pp. 393–410, 2021.

[10] F. Kong, Y. Li, H. Nassif, T. Fiez, R. Henao, and S. Chakrabarti, "Neural insights for digital marketing content design," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 4320–4332.

[11] M. Sarkar and A. De Bruyn, "LSTM response models for direct marketing analytics: Replacing feature engineering with deep learning," *Journal of Interactive Marketing*, vol. 53, no. 1, pp. 80–95, 2021.

[12] S. Tewari, P. Pathak, and P. Stynes, "A novel aspect-based deep learning framework (ADLF) to improve customer experience," in *Big Data Analytics*, S. N. Srirama, J. C.-W. Lin, R. Bhatnagar, S. Agarwal, and P. K. Reddy, Eds. Springer International Publishing, 2021, pp. 119–130.

[13] P. Manjula, N. Kumar, and A. A. Al-Absi, "Customer sentiment analysis using cloud app and machine learning model," in *Proceedings of International Conference on Smart Computing and Cyber Security*, P. K. Pattnaik, M. Sain, A. A. Al-Absi, and P. Kumar, Eds. Springer, 2021, pp. 325–336.

[14] M. Torrens and A. Tabakovic, "A banking platform to leverage data driven marketing with machine learning," *Entropy*, vol. 24, no. 3, 2022.

[15] S. Gutnik, "Application of data mining and machine learning methods to enhance the effectiveness of digital marketing strategies," in *Digital Strategies in a Global Market: Navigating the Fourth Industrial Revolution*, N. Konina, Ed. Springer International Publishing, 2021, pp. 131–144.

[16] A. Miklosik, M. Kuchta, N. Evans, and S. Zak, "Towards the adoption of machine learning-based analytical tools in digital marketing," *IEEE Access*, vol. 7, pp. 85 705–85 718, 2019.

[17] K. Jin, Z. Z. Zhong, and E. Y. Zhao, "Sustainable digital marketing under big data: An AI random forest model approach," *IEEE Transactions on Engineering Management*, vol. 71, pp. 3566–3579, 2024.

[18] T. Wang, C. He, F. Jin, and Y. J. Hu, "Evaluating the effectiveness of marketing campaigns for malls using a novel interpretable machine learning model," *Information Systems Research*, vol. 33, no. 2, pp. 659–677, 2022.

[19] E. Kongar and O. Adebayo, "Impact of social media marketing on business performance: A hybrid performance measurement approach using data analytics and

machine learning," *IEEE Engineering Management Review*, vol. 49, no. 1, pp. 133–147, 2021.

[20] R. Tidhar and K. M. Eisenhardt, "Get rich or die trying… finding revenue model fit using machine learning and multiple cases," *Strategic Management Journal*, vol. 41, no. 7, pp. 1245–1273, 2020.

[21] B. He, Y. Weng, X. Tang, Z. Cui, Z. Sun, L. Chen, X. He, and C. Ma, "Rankability-enhanced revenue uplift modeling framework for online marketing," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 5093–5104.

[22] S.-H. Lee, S.-H. Yoon, and H.-W. Kim, "Prediction of online video advertising inventory based on tv programs: A deep learning approach," *IEEE Access*, vol. 9, pp. 22 516–22 527, 2021.