Optimizing Marketing Strategies: Integration of Al-Biruni Earth Radius Algorithm for Feature Selection and Pipeline Regression Model

Khaled Sh. Gaber[1], Ahmed Mohamed Zaki[1], Marwa M. Eid[2,3,*], Doaa Sami Khafaga[4], Amel Ali Alhussan[4], Mahmoud Elshabrawy Mohamed[1]

[1]Computer Science and Intelligent Systems Research Center, Blacksburg 24060, Virginia, USA
[2]Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura 11152, Egypt
[3]Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura 35111, Egypt
[4]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

* Corresponding author: mmm@ieee.org

Emails khsherif@jcsis.org ,Azaki@jcsis.org; mmm@ieee.org; dskhafga@pnu.edu.sa; aaalhussan@pnu.edu.sa; mshabrawy@jcsis.org

Abstract

With the current business environment becoming increasingly ferocious, the effectiveness of digital marketing strategies is no longer a matter of debate as many organizations have realized the need to gain an edge over competition and improve the ROI with their marketing efforts. This study looks into the specifics of digital marketing effectiveness by, in the process, analyzing true indicators and key metrics. Demonstrating an understanding of the complexity of online marketing operations and the diversity of the variables involved, econometric techniques provide feature choice that affects campaign outcomes the most. At first, the variety of performance between different algorithms from feature selection gave the average error ranging from 0.38264 to 0.44194. However, following the optimization provides the tendency to see a decrease in mean errors and an improving performance. Afterward, the step of predictive modeling is implemented, employing diverse machine learning algorithms including ExtraTreesRegressor, GradientBoostingRegressor, SVR, and CatBoost to assess the effectiveness of foreshowing marketing outcomes. Before the optimization, the recommendations made by the predictive modeling are not too accurate and uniform for each algorithm. That being said, however, once the optimization is done, enhancement in prediction accuracy to the tune of substantial improvement is observed with metrics indicating the same as less MSE, RMSE, and R2. Contributing to a more thorough comprehension of the issue of selecting features and models for predicting as well as efficiency of digital marketing, the study also offers an understanding of the opportunities and obstacles that are present in the process of building digital marketing strategies. A thorough evaluation of top metrics and KPIs gives decision-makers data-driven tools to define their marketing activities, deliver tangible results, and stay relevant in the fast-paced digital environment of today.

Keywords: Digital marketing; Feature Selection; Predictive Modeling; Metrics; KPIs; Optimization.
MSC: 93B45; 83C20; 78M50

Doi : https://doi.org/10.21608/jaiep.2024.354005
Received: March 19, 2024; Accepted: April 22, 2024

1. Introduction

In digital marketing of today's day, online society has been developing so fast just to present an even wider environment of platforms and communication channels to businesses, the main task of which is to give good marketing. Today digital marketing can be seen as one of the most powerful tools for companies, who pursue to create a vivid and successful online brand, engage with their customers, and achieve their main goals. Yet, amid the vast ocean of digital data from these marketing operations, isolating quality insights that can help strategic decision moves remains an enduring hurdle.

The present paper is planned to explore the core parts of digital marketing analytics, which play an important role in providing a perspective on relying on data-driven techniques to strengthen marketing activities. At the core of this exploration lie several key areas that warrant attention and analysis: At the core of this exploration lie several key areas that warrant attention and analysis:

- Dataset Preprocessing: The process of data preprocessing commences with a transfer of raw data into a format that's ready for processing. This fundamental stage involves a bunch of actions, the popular of which consist of data cleaning, normalization, and feature engineering, which are must-haves if one wants the resulting dataset to be of high accuracy. By finely detailing the data, a broad and thorough base for a thorough analysis of modeling is made, and ultimately, more exact and worthwhile insights are derived.
- Descriptive Analysis Insights: Now, with our data set ready for study, it's time to dig deeper into the patterns and trends hidden in the data. Descriptive analysis methods like exploratory data analysis (EDA) show organizational leaders what important points can be used via visualizations like histograms, scatter plots, or correlation matrices. These representations carry the whole picture of the data along with the crucial details on the performance of the campaign, customer behavior, and so on that are essential for digital marketing performance.
- Performance Metrics Selection for Feature Selection: Feature choice is considered a critical component of performance metrics, which makes a direct impact on how we can assess the relationship between different algorithms and their effectiveness. Fitness rate is one of those that represent the average fitness and is another way used by organizations to identify the best algorithm for picking the variables that determine success in marketing.
- Performance Metrics Selection for Regression Models: To return the accuracy and cost-effectiveness of the regression model, the selection of the right performance metrics must be taken into consideration. An example of the role of metrics is in the estimation of the mean square error [MSE], the root means square error [RMSE], and the correlation coefficients [R] which in turn are indicative of the predictive ability of different algorithms. Better models are then chosen for marketing outcome forecasts.
- Results and Model Comparison: Lastly, this paper attempts to give the results of my research, Undoubtedly, this survey should include some type of evaluation (perhaps with the use of a comparison of different models) Performance analysis of regression models, weighing both accuracy, efficiency and their scalability enables organizations to take action on the underlying components of these models that are either strengths or weaknesses. Besides, this management gives them a chance to decide according to the necessary factors and it helps organizations to fine-tune their digital marketing strategies effectively, Data-driven insights enable organizations to optimize their digital marketing strategies to meet new levels of efficacy and cost-efficiency. In other words, through handling such information an organization acquires knowledge and means that can contribute to the success of digital marketing.

Literature review on digital marketing metrics and KPIs will be observed, the methodology according to which the dataset is going to be analyzed will be described and the results of the comprehensive analysis will be presented. Using this initiative, we seek to earn businesses important insights for planning, designing, and assessing the effectiveness of their activities, which finally leads to a reduction in marketing costs and increase in ROI, and thus the discussion on data-oriented marketing approaches in digital marketing will continue.

2. Literature Review

[1] With fluctuations in the market and competition as a phenomenon, the management and business owners couldn't simply sit tight and wait for their company to realize its growth and profitability goals. Business Intelligence [BI] is the primary tool that facilitates the improvement of the quality and

Doi : https://doi.org/10.21608/jaiep.2024.354005
Received: March 19, 2024 Accepted: April 22, 2024

timeliness of input-decision-making factors and adds valuable data analysis abilities to the decision-making process, allowing planners and top managers to make strategic partisan decisions. The benefits of BI regarding producing insightful and managerial-orientated reports that guide toward operational efficiency, lucrative opportunities, and strategic advantages are well established. On the flip side, questions regarding the effectiveness of BI tools on the quality of decisions and business growth remain topical. Different empirical aspects of AI applications foster the use of ML models that shed light on how ML models like the effect of AI systems on critical metrics like revenue development and customer behavior can be predicted. Integrating historical data into the multiple input elements through the process of machine learning, businesses become better equipped to make prudent decisions regarding the purchase of new tools and systems.[2] Neural networks as an emerging trend in predictive analytics in recent years. A real-world example of furniture sales prediction is illustrated using a public data set of historical sales data for a retail outlet. A series of forecasting models, ranging from traditional time-series techniques like Seasonal Autoregressive Integrated Moving Average [SARIMA] and Triple Exponential Smoothing to advanced methods like Prophet, Long Short-Term Memory [LSTM], and Convolutional Neural Network [CNN] are implemented. An illustration would use metrics and give the [RMS]and [MAPE]. It graduated the best performance resulting in the Stacked LSTM method. Moreover, findings indicate that the era prediction of the two models is perfect.[3] In 2015, consumers spent approximately $261 billion on products worth returning, and the percentage of goods returned from online channels was about 30%. Operations supervisors in product returns stand faced with the challenge of forecasting the volume of returns. On the other hand, manufacturers and retailers anticipate this. Data-oriented models for predicting the returns at the levels of line, independent, and franchise stores are built and validated using the varied, yet enriched on the product and retailer level, dataset. The implementation of the [LASSO] model, which achieves better accuracy than all other models by choosing only the informative terms from a multitude of combinations is the key factor for achieving optimization. [3] offers a valuable predictive framework for manufacturers to track product returns.[4] The present research in this field is not so adequate. Bridging the identified gap, the concept that consists of sources of algorithmic biases in marketing is given, applying the micro foundations theories of the dynamic's capabilities. A comprehensive literature review with a focus on relevant subdimensions is conducted and in-depth interviews are conducted to determine the different dimensions and related subdimensions of algorithmic bias. By introducing a dynamic algorithm governing framework, [4] tries to reduce biases in ML-based marketing decision-making [3]. Customers could be detected by the outcomes of predicted processes which enable customers to have a better experience. This would allow service level to be improved and customers to be more satisfied. A blended system of supervised/unsupervised machine learning methodologies is used to generate a customer satisfaction prediction model which can address noisy and high-variance data. By using methods such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for detection of the nearest neighbor and oversampling for dealing with imbalanced classification, the proposed technique gets high accuracy in forecasting customer dissatisfaction.[6] However, it should be intellectually acknowledged that Data Sciences' development in digital marketing implies the development of better management practices. A critically inclusive literature review provides both traditional analysis to fresh analysis tools as well as Data science performance indicators in digital marketing. Insight gleaned from the review serves to improve the development of innovative techniques in Data Mining & the Process of knowledge discovery which are aimed at coming up with advice that may be used to enhance the digital marketing strategy [7]. It is thus important for content marketing specialists to be able to comprehend the ML applications. In one study, three of the ML techniques for automatic tagging and categorization of news articles available online were compared, and the Neural Network showed the greatest performance, suggesting its suitability for cross-platform applicability and content performance analysis.[8]A data-driven review of recent achievements in resolving the difficult problem of uncertainty known as optimization is presented pointing out main challenges and possible solutions. [8] review should be conducted by comprehensively analyzing relevant publications targeted to shed light on integrating machine learning and mathematical programming as a solution for decision-making under uncertainty.[9] Tracking customer needs eliciting and monitoring is the key point for business comprising customer-centered products and services. The possibility of solving those tasks based on the prioritization and quantification of social media data from customers' speaking is shown in this sentence. Through applying a supervised learning approach, classification models can be developed to calculate the odds of tweets with all needs defined uplifting the ability to demand differentiation and monitoring.[10] With deep learning in the scene of business analytics, the field is witnessing a paradigm shift. Nonetheless, there is a huge variance regarding the application of deep learning to analytics in business. Researching deep learning for business analytics is a vital approach that is reviewed here focusing on its operational values, and applications with potential in the business world. The practical applications do point, however, to a new generation of deep learning architectures with a truly unique character of purposeful work as the essence of their value.[11] A benchmark study combined with an analysis of machine learning algorithms is presented, exploring the market strategies development in a digital economy that offers strategies specific to machine learning and its benefits implementation on the market strategies.

3. Proposed Methodology

As Shown in figure 1 the analysis of the digital marketing data through the machine learning techniques that are about to be addressed in the paper. The framework for data collection starting with the dataset procurement, feature engineering, and model evaluation is depicted sequentially from the acquisition of the dataset until the model selection of the most suitable machine learning model. Initially, data collection is done including the successive period covering the steps of feature selection, data preprocessing, application of regression models, obtaining performance results, and therefore, initial model selection in the process. This visual representation presents the systematized methodology of the paper for the effective implementation of machine learning in digital marketing analysis through the following diagram.

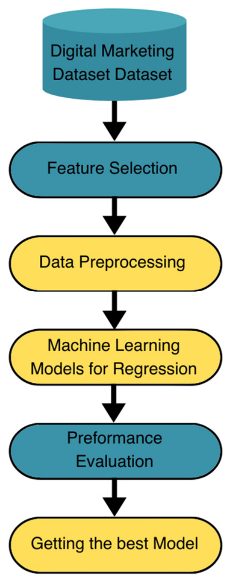

Figure 1: Proposed Methodology of This Paper

3.1 Dataset Description

The Dataset [12] Description section provides a detailed description of the digital marketing dataset used in this study. It is a complex set of different columns that contain different sets of data all to measure and enable the performance metrics of a campaign.

- id: One of the main features of the dataset is the unique identifier for each entry that simplifies data management and helps analytics.
- c_date: This run shows the date the marketing campaign was launched and consequently places a temporal dimension for analysis.
- campaign_name: Descriptive labels are assigned to each campaign making the identification and classification process easier.
- category: Indicates the type of each campaign or the type or category of each campaign like social media, search engine, influencer marketing, or traditional media.
- campaign_id: Together with the entry ID, this column will assign a unique identifier for each campaign so that there is no overlapping and it is easy to follow.
- impressions: The metric shows the total number of impressions (an impression is a single time when the content of a campaign is displayed to users) produced by each campaign.

Doi : https://doi.org/10.21608/jaiep.2024.354005
Received: March 19, 2024 Accepted: April 22, 2024

- mark_spent: It reflects the monetary amount invested in each campaign, and therefore helps understand the financial investment needed for the campaign execution.

- clicks: Shows the number of times when each campaign was clicked, and reflects how much engagement and emotional connection the campaign content has with the user.

- leads: Counts the number of leads which refers to the number of potential customers who have shown interest in the product or service provided.

- orders: Shows the number of orders made because of the campaigns, examining the campaign's effectiveness in both closing leads and bringing sales.

- revenue: The main point is the total revenue, and it is the result of the campaign activities operations, which reflect the financial returns or profitability.

This data set covers various types of marketing campaigns targeted and also includes their parameters and customer engagement characteristics generated from these campaigns. This analysis helps to derive insights into what is working, and what is not, and to measure key communications variables for budget allocation and marketing strategy optimization.

Figure 2 comprises two boxplot charts: "Customer Acquisition Cost (CAC) by Campaign " and " Cost Per Click (CPC) by Campaign." The CAC chart which depicts the dispersion of CAC for each, advertising campaign including the range of each campaign scale (per CPC) is shown in a box plot. While the best campaigns with high ROI include such examples as advertisements on Facebook Ads and Influencer Campaigns, we can properly evaluate the spending efficiency and the effectiveness of the campaign through such exceptional CAC values. The diagram depicts the CPC distribution throughout the advertising campaigns - the chart demonstrates the volatility and celebrates the rare spikes, particularly in campaigns like "Instagram Ad 1." This capacity of the diagram helps with the budget allocation optimization and campaign efficiency analysis.

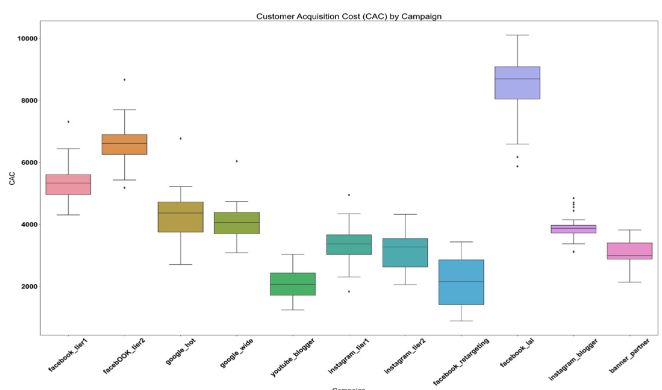

Figure 2: A Box Plot for Comparison of Customer Acquisition Cost (CAC)

3.2 Descriptive Analysis

The Statistical Analysis Section is constructed to be a strong base in the maze of digital marketing data by applying a variety of statistical measures and visualizations that will showcase the primary insights. The section exploring the means, medians, modes, and quantiles of the most important metrics such as revenue, clicks, and impressions reveals the performance-driven trends and patterns that underlie the campaign performance. By the means of carefully employing bar charts, graphs, and boxplots, marketers, take into account all distributional characteristics and variability of their data. From the exploration of the revenue generation collective across multiple campaign categories to the analysis of the clicks and impressions, these visual representations are of paramount importance as they provide interesting and detailed context and depth. They work as compasses for strategic decisions that empower marketers to allocate resources properly and fine-tune marketing strategies for effective impact and high return on investment. In short, the Descriptions Analyses Section works as a lighthouse shining on through the data-driven marketing competence and thus brings informed business management in the ever-changing digital environment.

The mixed data Figure 3 displays three bar diagrams that give an overview of the revenue, clicks, and impressions across various marketing camps and genres. Every bar is representative of the campaign which is manually labeled according to the campaign being run on the four marketing channels—Social, Search, Influencer, and Display. Thus, any given bar can be easily interpreted from the performance metrics. The visualization of the data demonstrates the three understandings of revenue contributions, user engagement metrics, and audience reach which outlines that specific marketing initiatives may bring financial returns, audience interaction, and wider audience reach, respectively. Such a detailed visualization becomes an effective strategic tool for marketers which lets them track the effectiveness of their marketing strategy and adjust marketing ads to deliver the maximum touchpoint in investment.

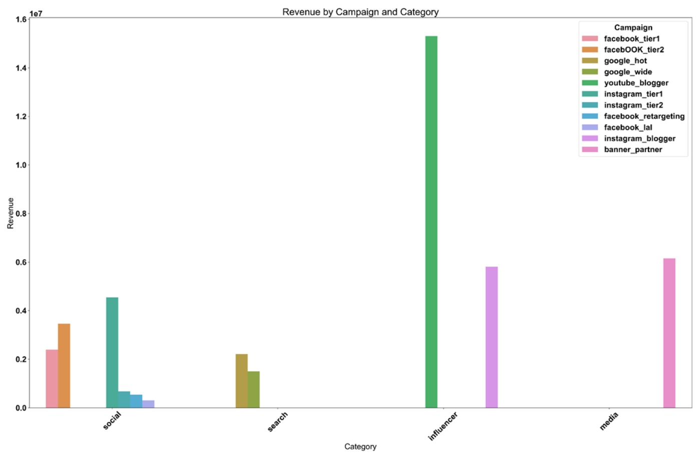

Figure 3: Barplot Comprehensive Analysis of Marketing Campaign Performance Across Categories.

The overall figure, which is abbreviated as Figure 4, provides three different scatter plots that altogether showcase vital relations between campaign performance. The first scatter plot, titled "Revenue vs. Orders," aims to discover a correlation between revenue and orders received by a business. The next plot, entitled "Cost Efficiency Analysis", shows the correlation of the CPC (Cost Per Click), CAC (Customer Acquisition Cost), and the number of products sold. Lastly, our "Cost vs. Revenue" plot shows that positive traction in sales is achieved through marketing expenditure. As a whole, these visualizations present the important analysis surrounding the campaign effectiveness, cost efficiency, and revenue gain to recognize the key design of marketing resources.

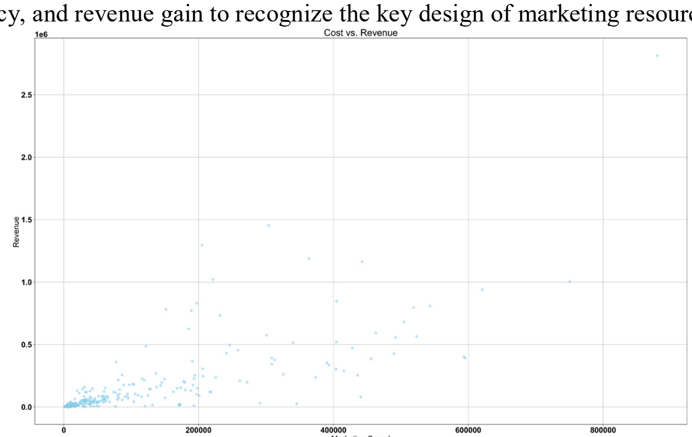

Figure 4: Scatter Plot Exploring Relationships in Marketing Campaign Performance Revenue

A correlation heatmap is a pictorially entitled visual method used to show the relationship between metrics factors such as impressions, clicks, revenues, CPC, CAC, and ROMI and it uses color graded for correlations' strength representation. This is represented by each cell with a given correlation coefficient. The metric technique is shown on the left side of the graph with the correlation coefficients running on the right bank. And so positive correlation compares a small relationship to a moderate relationship, whereas a negative correlation reveals it has an inverse contribution.

Doi : https://doi.org/10.21608/jaiep.2024.354005
Received: March 19, 2024 Accepted: April 22, 2024

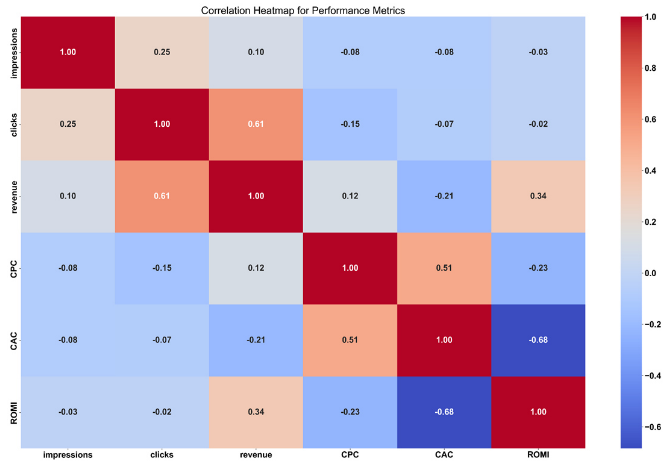

Figure 5: A Correlation Heatmap of Marketing Performance Metrics

3.3 Feature Selection

Feature selection is one of the main things considered in the machine learning process and its goal is to determine the minimal set of features with the score data not being adversely affected. The model performance and its computational efficiency are modernized by noise reduction and terms elimination (FS) [13] The main tasks performed by it are dimensionality reduction, accuracy enhancement, and data visualization. Some of the FS algorithmic evaluators employ measures of evaluation and search strategies. Nevertheless, evolutionary computation algorithms, such as genetic algorithms and particle swarm optimization, which are the most commonly applied methods for dimension reduction, are usually very consuming in terms of computational resources.

3.3.1 Feature Selection Models

1. **Al-Biruni Earth Radius (BER) Optimization Algorithm:**
   - The idea of the Al-Biruni Earth Radius optimization (BER) algorithm is based on the method that Al-Biruni used to calculate the radius of the Earth. It performs the task of generating the optimal solution and this solution is supposed to be within the constraints which is represented by each element of the population vector in the BER algorithm. The problem is that the population is evaluated by the fitness function F and we look for the vector S* with the maximal fitness. The first optimization steps depend on several components, like the specification of the fitness function, size limits, population size, dimension, and number of solutions. BER algorithm is deployed by employing exploration operation and exploitation operation seeking to optimize the search of the process.
2. **Grey Wolf Optimization (GWO) Algorithm:**
   - [14], GWO is a recent innovation based on the concept that swarms can solve complex problems more efficiently than individual elements. The GWO welcomes the behaviors of grey wolf packs into its social factors such as the organization and the hunt. The hierarchy is copied from such groups. In the pack, the 'alpha' Wolf is the one who is at the highest rank and he controls the group while hunting, eating, and migrating. If the alpha wolf is missing, the beta one comes forward to lead the pack. However, omega and delta, which are also the two core members of the pack, but are not that influential, have a lot to contribute to the pack. Hence built up the structure of the GWO algorithm.
3. **Grey Wolf Optimization (GWO) with Particle Swarm Optimization (PSO):**
   - At the same time, the authors Al-Tashi and others proposed a hybrid approach that is based on the grey wolf optimization (GWO) and particle swarm optimization (PSO) procedures for the feature selection.[14]
4. **Particle Swarm Optimization (PSO):**
   - Particle Swarm Optimization (PSO) designed by Kenney and Ebenhart in 1995 has been widely popular among the feature selection methods because it imitates the way birds in flocking create a social system for solving optimization problems [15]. An additional characteristic of heuristic optimization enabled by swarm intelligence is PSO's computation efficiency and fast convergence. The PSO technique is simulated within the system by swarming particles that each have a velocity to move and a position where they are located. The particle gains a new understanding of its own space, its own experience, and that of its neighbor. Then, it adjusts its position and velocity. The term personal best is used to refer to the one particle's best position formerly, while the very best position of

the entire population of particles is marked as global best. Thus, these particles´ velocity, as well as position, are transformed, always seeking the best solution considering both personal and global conditions. Therefore, the algorithm terminates after the standard attainment of a criterion that could be any of the best fitness values, a specific number of iterations as defined by the system.

5. **Whale Optimization Algorithm (WOA):**
   - WOA represents the collective organization as an intelligent swarming algorithm for continuous optimization. Recent results have revealed that their method outperforms other meta-heuristics algorithms in terms of performance [15]. First of all, it indicates that this one algorithm has fine peculiarities, such as the simplicity of its realization and solidness, that different nature-spired algorithms do not have. The algorithm has also no need for the mean controllers; time interval is the only parameter that is essential for adjustment.
6. **population-based algorithm (BBO):**
   - BBO is an evolutionary-based genetic algorithm developed by Dan Simon [17], which is referred to as the BBO. Motivated by the natural dynamic of biogeography, which involves species being bound to their natural habitat, the method models candidate solutions as habitats called species. Every focus habitat has particular attributes like rainfall and the size of the land. A population of BBO is set up and consists of several habitats which are generated randomly. The habitat is assessed based on its Habitat Suitability Index (HSI) which shows a site's propensity for a specific solution.
7. **The multi-verse optimizer (MVO):**
   - MVO mimics the physics theory of multi-verse,[18] in which a system of competing universes is carried out, and at the end of the process, sustainability is achieved. Mjirjalili introduced it to help solve problems in the optimization field.
8. **Stain Bowel Obstruction (SBO):**
   - In a similar way to other metaheuristic algorithms, the SBO algorithm initiates with the generation of an initial population of bowers.[19] This initial population comprises a vector articulating the positions of the nodes so that an optimum solution can be derived. Such vertices begin at the origin, are randomly distributed, limited at the top and bottom with upper and lower vertex values. The interpretation of bower values has a direct representation of the optimization variables since they work towards the aesthetics of the bower.
9. **A wrapper-based Genetic Algorithm (GA):**
   - [20] in wrapper-based Genetic Algorithm (GA) is used to choose the high-powered features. The wrappers within the GA deduce whether the subsets are fit to be passed using some test and evaluating each one of them by using a model. The GA engineers a search algorithm to create a population of candidate solutions (phenotype), and it constructs an optimization problem for the species hence a due solution. Usually, a candidate solution as the name itself reveals can be linked to a component of the solution for a specific problem. Individual genotype, or chromosome, is a descriptive unit of a candidate solution. These are the set of generated parameters that a GA produces. Like, the population itself is a cluster of all participants.
10. **the Firefly Algorithm (FA) Algorithm:**
    - The main purpose of the Firefly Algorithm (FA) is essentially to find patterns in the observation variables and simultaneously define classes that each variable belongs to while minimizing the correlation between the classes [21]. Every pair of variables, a public factor, which are essential to the target variable, belongs to a particular category of variables. The optimal subset of each variable is identified using factors analysis for each variable and all of its combinations through the system are presented under the results section of the study based on subsets information.

3.3.2 Performance Metrics for Feature Selection

Feature selection in the area of machine learning and data mining is one of the key elements to obtaining a precise and proficient model for the task. We select evaluation metrics for the status of feature selection algorithms in various modes. This is where we will investigate core metrics such as an average error, the purity of selected genes, fitness, and many others. The knowledge of the indicator is significant in the evaluation and choice of the methods for data selection, thus more information enhances the performance of the model and data-driven decisions in the real world [21].

Table 1: Criteria for evaluating Feature Selection result.

Doi : https://doi.org/10.21608/jaiep.2024.354005
Received: March 19, 2024 Accepted: April 22, 2024

Metric	Formula
Average Error (AE)	$AE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)$
Average Select Size (ASS)	$ASS = \frac{1}{n}\sum_{i=1}^{n} s_i$
Average Fitness (AF)	$AF = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$
Best Fitness (BF)	$BF = \min(f(x_1), f(x_2), \dots, f(x_n))$
Worst Fitness (WF)	$WF = \max(f(x_1), f(x_2), \dots, f(x_n))$
Standard Deviation Fitness (SDF)	$SDF = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(f(x_i) - AF)^2}$

The following metrics are important parts of our full review framework: So, in this case, it is utterly irresponsible not to acknowledge the fact that as much as one can nullify a possible preterm birth.

- **Average Error:** The average error in matrices is the arithmetic mean of the differences between the predicted values and the true value across a given data set. This metric is used as a measure of the overall model predictiveness showing the level of coincidence between the predicted and actual values on average.
- **Average Select Size**: In a matrix context, the 'average select size' represents a random variable with the mean value equal to the dimensionality or the magnitude of elements or submatrices being selected in comparison to the total matrix size. This factor is critical in various areas, especially optimization and extraction algorithms where subsets of data or features are chosen based on these criteria.
- **Average Fitness:** The term average fitness in connection with the matrix is generally referred to as a mean performance score possessed by a population of organisms, i.e. matrices, during evolutionary computation or genetic algorithm. The fitness metric is determined using a predefined fitness function. This provides a yardstick to measure the population's aggregated performance and an indicator for the effectiveness of the total optimization.
- **Best Fitness:** This population's best fitness means the goal score attained by every individual in the population, which shows the highest attainable performance in optimization and evolutionary computation. This metric identifies those individuals who are more adept among the population.
- **Worst Fitness:** Alternatively, the lowest fitness in the matrix model indicates the lowest performance score of a given population, which is the mathematically worst performance achieved during an optimization or evolutionary computation process. The purpose of this metric is to determine the individual with the lowest score from amongst the whole population.
- **Standard Deviation Fitness**: this standard deviation of fitness values is the measure of how much the fitness scores vary from each other, across the whole population. This dispersion metric stands as an indicator of the extent to which the fitness value differs from the mean fitness of the population, any such differences are due to diversity or spreading of performance within the population, consequently, this information is vital in the determination of population convergence or diversity maintenance strategies

3.4 Machine Learning Regression

Concerning this part, we move into the aspect of using regression models to estimate student performance. Regression methods which include the utilization of various regression techniques are the key instrument that shall be used to discover the complex patterns hidden in our dataset. The fallacy of our predictive loop is weighed using the key performance metrics, such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), etc. to assess the predictive accuracy of the different trainings, and as such, provide us with valuable insights into the effectiveness of the predictive approach.

3.4.1 Regression Models

1. **Pipeline:**

The pipeline is a concept in sci-kit-learn for streamlining a lot of the routine processes, including combining different models. The mathematical equation is not directly applicable here as it's more of a software engineering concept.

2. **Extra Trees Regressor:**

Extra Trees algorithm,[22] the development of Random Forest has addressed overfitting by expanding to reduce variability. It operates like the Random Forest but it differs by selecting the best feature for the split randomly and differentiating each regression tree by the whole training dataset, not by the bootstrap replicas.

3. **Gradient Boosting Regressor:**

Gradient Boosting (GB) is an algorithm that creates an ensemble of decision trees progressively to obtain precise predictions [23]. It begins from an initially poor model whose first tree in the ensemble is shallow and later adds additional trees. Subsequently, each tree tries to fix the mistakes of the previous tree by way of reduction of loss function via gradient descent. The GB pools together the forecasts of numerous weak models and produces a single, final prediction.

$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i)$, where f_k is the k-th weak learner.

4. **Support Vector Regression (SVR):**

The SVM (Support Vector Machine) is one of the most commonly used computational intelligence techniques that has been implemented in predictive modeling for building energy and renewable energy generation. It provides a parsimonious solution structure and deals satisfactorily with nonlinear situations even when the number of training data samples is low [22].

$\hat{y}_i = \sum_{j=1}^{n} (\alpha_j - \alpha_j')K(x_i, x_j) + b$, where K is the kernel function, α and α' are the Lagrange multipliers, and b is a bias term.

5. **CatBoost:**

CatBoost is similar to XGBoost but is designed to handle categorical features efficiently [24]. The prediction is given by:

$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i)$, where f_k is the k-th weak learner.

6. **Random Forest Regressor:**

Random Forest Regressor combines the predictions from multiple decision trees [25]. The prediction is an average of the predictions from all the trees.

7. **Decision Tree Regressor:**

Decision tree (DT) is an effective technique for both classification and regression issues [22]. The essential idea behind the decision tree algorithm is to break down a complex problem into some simpler parts, which ideally result in a solution that is easier to understand.

8. **XGBoost:**

XGB is a highly popular ML algorithm that performs very well in many application areas [23]. It gradually constitutes an assembly of trees, with each tree learning from the mistakes of the preceding ones. The algorithm uses the technique of regularization including tree pruning (cutting off tree branches (subtrees)) and column subsampling to regulate overfitting and provide better generalization.

$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i)$, where f_k is the k-th weak learner.

9. **Linear Regression:**

Linear Regression predicts the output as a linear combination of input features [22]. The prediction is given by:

$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_n x_{in}$, where $b_0, b_1, \ldots, b_n$ are the coefficients.

10. **MLP Regressor (Multi-Layer Perceptron):**

The prediction in a Multi-Layer Perceptron is given by [26]:

Doi : https://doi.org/10.21608/jaiep.2024.354005
Received: March 19, 2024 Accepted: April 22, 2024

$y\hat{}_i = \phi(x_i) = f^{(L)}(f^{(L-1)}(\ldots f^{(1)}(x_i;\theta^{(1)})\ldots;\theta^{(L-1)});\theta^{(L)})$, where $f^{(l)}$ is the activation function in layer l and $\theta^{(l)}$ are the weights.

11. **K-Neighbors Regressor:**

The KNN regression employs a relatively simplistic technique of finding the K nearest neighbors (K represents those training data points that differ from each other the least on the feature space or known data points similar to the new observation the least) from the past training dataset to predict the value of a new observation [26].

$y\hat{}_i = \frac{1}{k}\sum_{j=1}^{k} \quad y_{N_j}$, where N_j is the j-th nearest neighbor.

3.4.2 Performance Metrics for Regression

This section summarizes some of the fundamental performance metrics employed in assessing the effectiveness of regression models [27]. The metrics are (MSE), (RMSE), (MAE), (r), (R2), (NSE), and (WI). They measure the accuracy, the precision, and the reliability of the regression model.

Table 2: Criteria for evaluating regression result.

Metric	Formula
MSE	$\frac{1}{N}\sum_{n=1}^{N}(y_i - \grave{y_i})$
RMSE	$\sqrt{\frac{1}{N}\sum_{n=1}^{N}(V\hat{}_n - V_n)^2}$
MAE	$\frac{1}{N}\sum_{n=1}^{N}\lvert V\hat{}_n - V_n\rvert$
NSE	$1 - \frac{\sum_{n=1}^{N}(V_n - V\hat{}_n)^2}{\sum_{n=1}^{N}(V_n - V\hat{}_n)^2}$
WI	$1 - \frac{\sum_{n=1}^{N}\lvert V\hat{}_n - V_n\rvert}{\sum_{n=1}^{N}\lvert V_n - V^{-}_n\rvert + \lvert V\hat{}_n - V^{-}_n\rvert}$
R^2	$1 - \frac{\sum_{n=1}^{N}(V_n - V\hat{}_n)^2}{\sum_{n=1}^{N}(\sum_{n=1}^{N}V_n) - V_n)^2}$
r	$\frac{\sum_{n=1}^{N}(V\hat{}_n - V\hat{}_n)(V_n - V^{-}_n)}{\sqrt{\left(\sum_{n=1}^{N}\left(V\hat{}_n - \underline{V\hat{}_n}\right)^2\right)(\sum_{n=1}^{N}(V_n - V^{-}_n)}}$

- **Mean Squared Error (MSE):** This quantity is defined as the mean square difference between the predicted and the observed values which is used as a metric for the model accuracy. Lower MSE indicates good model performance.
- **Root Mean Squared Error (RMSE):** RMSE is the square root of MSE, so it represents the average magnitude of errors. Low RMSE is better than MSE, likewise.
- **Mean Absolute Error (MAE):** MAE measures the average absolute difference between observed and estimated values, thus, indicating the error size range.
- **Mean Bias Error (MBE)**: MBE is the mean squared difference between predicted and observed values, which indicates overestimation for positive values and underestimation for negative values.
- **R (Correlation Coefficient):** This coefficient indicates the strength and direction of a linear relationship between variables, where values close to 1 show the existence of a strong positive correlation.
- **R² (Coefficient of Determination):** R² is the proportion of the dependent variable variance that could be explained by the independent variable. Higher R² values imply better model fitting.
- **RRMSE (Relative Root Mean Squared Error):** RRMSE is a measure of normalized prediction accuracy, which takes both the smallest and largest values.
- **NSE (Nash-Sutcliffe Efficiency):** NSE compares the accuracy of model predictions with the mean value of the actual metrics. Values closer to 1 suggest better model performance.
- **WI (Willmott Index):** WI measures the concordance of the observed and the model estimated values, with WI = 1 meaning the perfect agreement.
- **Fitted Time:** This shows the time that has elapsed to either train or fit the model to the data, where the smaller value denotes a faster processing time.

4 Results

This section is the report of the feature selection results as well as regression models. The most important thing in regression models is feature selection which is to determine only the most relevant predictors.

4.1 Comparative Performance Metrics:

The results of feature selection are showcased in Table 3, highlighting the performance metrics across various feature selection techniques.

Table 3: Feature Selection Result.

	bBER	bGWO	bGWO_PSO	bPSO	bWAO	bBBO	bMVO	bSBO	bFA	bGA
Average error	0.38264	0.39984	0.43914	0.43364	0.43344	0.40184	0.41034	0.44194	0.43204	0.41344
Average Select size	0.33544	0.53544	0.66874	0.53544	0.69884	0.69924	0.63194	0.70574	0.56994	0.47784
Average Fitness	0.44584	0.46204	0.47034	0.46044	0.46824	0.46614	0.49014	0.50014	0.51234	0.47344
Best Fitness	0.34764	0.38234	0.42384	0.44074	0.43234	0.45584	0.41534	0.44324	0.43104	0.37674
Worst Fitness	0.44614	0.44924	0.53384	0.50844	0.50844	0.54234	0.53334	0.52294	0.52864	0.49184
Standard deviation Fitness	0.26814	0.27284	0.29104	0.27224	0.27444	0.31714	0.32294	0.33314	0.30904	0.27444

The performance of different regression models is shown in Table 4. The metrics for assessing performance include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), correlation coefficient (r), Coefficient of Determination (R^2), Nash-Sutcliffe Efficiency (NSE), Willmott's Index (WI), and Fitted Time (FT)These measures give the information about the correctness, precision, and computational efficiency of each regression model.

Table 4: Regression Result.

Models	mse	rmse	mae	r	R2	NSE	WI	Fitted Time
pipeline	8.85E-32	2.97E-16	1.92E-16	1	1	1	1	0.171702
ExtraTreesRegressor	4.14E-06	0.002036	0.000714	0.997612	0.99523	0.988482	0.965987	0
GradientBoostingRegressor	9.28E-06	0.003047	0.001672	0.992541	0.985137	0.974194	0.920367	0.004001
SVR	1.22E-05	0.003499	0.001834	0.983553	0.967376	0.96597	0.912646	0.360941
CatBoost	1.98E-05	0.004446	0.002595	0.978522	0.957506	0.945055	0.876421	77.90268
RandomForestRegressor	2.69E-05	0.005185	0.001534	0.980812	0.961993	0.925256	0.926933	0.863512
DecisionTreeRegressor	5.01E-05	0.007079	0.00238	0.948931	0.900469	0.860696	0.886641	0.258784
XGBoost	0.000138	0.011767	0.009797	0.960949	0.923422	0.615115	0.533362	31.0902
LinearRegression	0.000145	0.012053	0.009602	0.882677	0.779119	0.596156	0.542641	0.019879
MLPRegressor	0.0002	0.014153	0.00936	0.824396	0.679629	0.443231	0.554197	5.864344
KNeighborsRegressor	0.00029	0.017043	0.010591	0.616593	0.380187	0.192628	0.495539	0.025077

Referring to Figure 6, we can see a radar plot showing the performance of different regression models for exam score prediction. This visualization includes indicators such as MSE and RMSE which makes the overall picture clear. While the plot swiftly demonstrates the strengths and weaknesses of the models, the overall performance is depicted in a precise, summarized manner.

Doi : https://doi.org/10.21608/jaiep.2024.354005
Received: March 19, 2024 Accepted: April 22, 2024

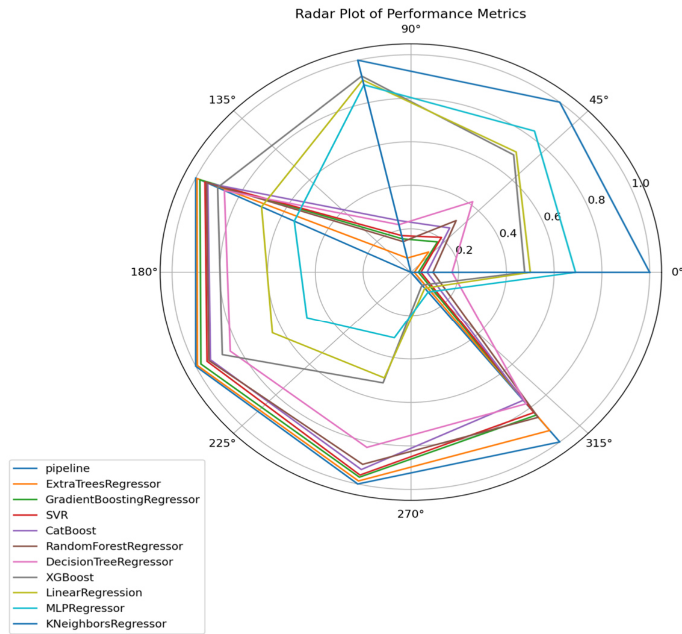

Figure 6: Radar Plot of the Performance Metrics

4.2 Discussion of Findings:

The evaluation of the models involved in the prediction was held and some significant performance was observed in the different regression models.

Impressive results were achieved by various metrics, especially in models like ExtraTreesRegressor and GradientBoostingRegressor. A high prediction accuracy was recorded, as the MSE, RMSE, and MAE indicators turned out to be very minimal. Also, strong correlations between the model predictions and the observations were visible as reflected by high values of R^2 and NSE.

Then again, a satisfactory degree of computational efficiency was maintained for all the models assessed. Despite some models presenting a higher Fitted Time, the overall computational charges were affordable for the accuracy level of foretelling.

Unlike others, some models show worse performance along various criteria besides that, these models provide ease of use and simplicity but struggle to match the accuracy and efficiency of competitors.

The results reinforce that a balance between the two criteria is necessary when choosing regression models for predictive tasks. By emphasizing these considerations, sound choices can be made for models that are best suited for the respective predictive objectives.

Examining the regression models in Figure (7) shows how strongly the regression models align with real-world performance by way of a Correlation heat map. The changing colors clearly show the relationship patterns that are used to assess the reliability of the predictive model. Explore together visual contexts to see a greater meaning in predictive modeling.

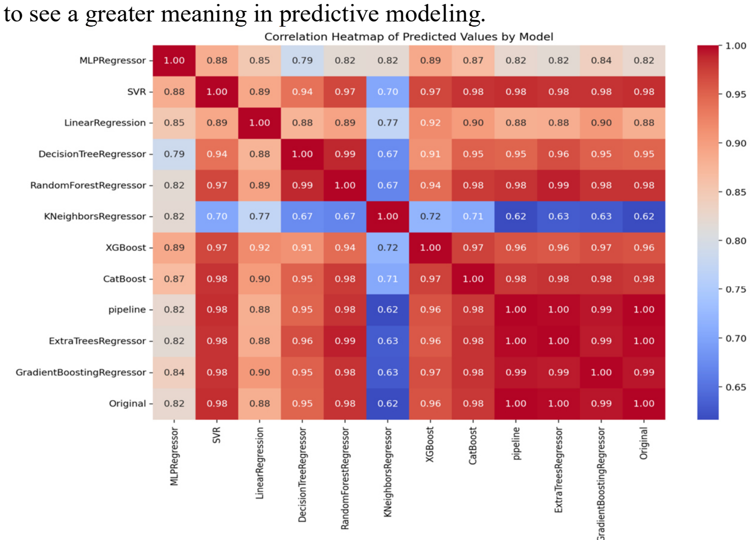

Figure 7: Correlation Heatmap

As we can see from Figure (8), where the comparisons of machine learning models on their Mean Squared Error(MSE) basis, it is clear that KNeighborsRegressor has the lowest error, meaning that its predictions are most accurate. On the other hand, the pipeline model's MSE level is the highest, showing it's the least accurate prediction approach.

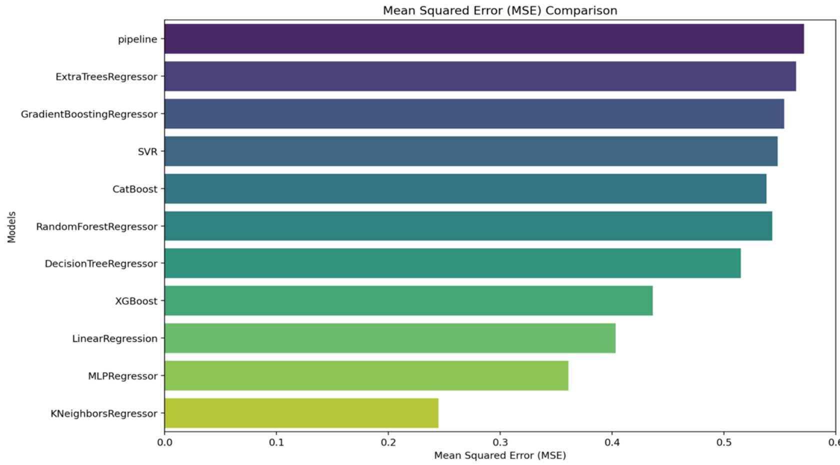

Figure 8: A Comparative Analysis of Regression Models

5. Conclusion

Lastly, our study has been very informative for us about how predictive modeling works in digital marketing analytics. Evaluating different kinds of regression models and feature selection methods has let us figure out the significant factors of predictive accuracy in the framework of big data for digital marketing.Our outcomes emphasize the necessity of being careful about the model selection and feature engineering to get fairly precise and reliable outputs for different marketing metrics. The examination of regression models has led to varied degrees of performance measures, underscoring the fact that a balanced consideration of predictive accuracy, computational effectiveness, and model interpretability needs to be done in the digital marketing realm. Furthermore, the examination of the usefulness of feature selection techniques also gave a clear picture of their value to the model's performance, as some were better than others in selecting relevant predictors and improving predictive power for marketing campaigns, customer behavior, and other important key measurements.Therefore, we can argue that a systematic way of doing predictive modeling in digital marketing is essential in choosing the model, engineering the features, and evaluating the performance, as summarized in the above-mentioned bullet points. Therefore, the findings of this research can be used by digital marketers and analysts to aid them in decision-making and optimize their digital marketing strategies to achieve business outcomes and return on investment. While moving forth, more investigation can be availed of involving better machine learning systems relevant to the peculiar roles of digital marketing analytics. Along with that, the longitudinal studies can focus on the long-term prospective validity of marketing models and their consequences for marketing strategy optimization and business success in the digital age. Hence, we offer a significant contribution to the field of predictive analytics in digital marketing and we suggest some practical directions for marketers, analysts, and decision-makers who are aiming to capitalize on data-driven approaches and increase the good results of their business activities in the dynamic digital environment.

Funding: "This research received no external funding"
Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Tripathi, M. A., Madhavi, K., Kandi, V. S. P., Nassa, V. K., Mallik, B., & Chakravarthi, M. K. (2023). Machine learning models for evaluating the benefits of business intelligence systems. The Journal of High Technology Management Research, 34(2), 100470. https://doi.org/10.1016/j.hitech.2023.100470

[2] Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. International Journal of Information Management Data Insights, 2(1), 100058. https://doi.org/10.1016/j.jjimei.2022.100058

[3] Cui, H., Rajagopalan, S., & Ward, A. R. (2020). Predicting product return volume using machine learning methods. European Journal of Operational Research, 281(3), 612–627. https://doi.org/10.1016/j.ejor.2019.05.046

[4] Akter, S., Dwivedi, Y. K., Sajib, S., Biswas, K., Bandara, R. J., & Michael, K. (2022). Algorithmic bias in machine learning-based marketing models. Journal of Business Research, 144, 201–216. https://doi.org/10.1016/j.jbusres.2022.01.083

[5] Akhavan, F., & Hassannayebi, E. (2024). A hybrid machine learning with process analytics for predicting customer experience in online insurance services industry. Decision Analytics Journal, 11, 100452. https://doi.org/10.1016/j.dajour.2024.100452

[6] Saura, J. R. (2021). Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics. Journal of Innovation & Knowledge, 6(2), 92–102. https://doi.org/10.1016/j.jik.2020.08.001

[7] Salminen, J., Yoganathan, V., Corporan, J., Jansen, B. J., & Jung, S.-G. (2019). Machine learning approach to auto-tagging online content for content marketing efficiency: A comparative analysis between methods and content type. Journal of Business Research, 101, 203–217. https://doi.org/10.1016/j.jbusres.2019.04.018

[8] Ning, C., & You, F. (2019). Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. Computers & Chemical Engineering, 125, 434–448. https://doi.org/10.1016/j.compchemeng.2019.03.034

[9] Kühl, N., Mühlthaler, M., & Goutier, M. (2020). Supporting customer-oriented marketing with artificial intelligence: Automatically quantifying customer needs from social media. Electronic Markets, 30(2), 351–367. https://doi.org/10.1007/s12525-019-00351-0

[10] Kraus, M., Feuerriegel, S., & Oztekin, A. (2020). Deep learning in business analytics and operations research: Models, applications and managerial implications. European Journal of Operational Research, 281(3), 628–641. https://doi.org/10.1016/j.ejor.2019.09.018

[11] Gutnik, S. (2021). Application of Data Mining and Machine Learning Methods to Enhance the Effectiveness of Digital Marketing Strategies. In N. Konina (Ed.), Digital Strategies in a Global Market: Navigating the Fourth Industrial Revolution (pp. 131–144). Springer International Publishing. https://doi.org/10.1007/978-3-030-58267-8_10

[12] Digital Marketing Metrics & KPIs to Measure (SQL). (n.d.). Retrieved May 5, 2024, from https://www.kaggle.com/datasets/sinderpreet/analyze-the-marketing-spending

[13] Tawhid, M. A., & Ibrahim, A. M. (2020). Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm. International Journal of Machine Learning and Cybernetics, 11(3), 573–602. https://doi.org/10.1007/s13042-019-00996-5

[14] Al-Tashi, Q., Md Rais, H., Abdulkadir, S. J., Mirjalili, S., & Alhussian, H. (2020). A Review of Grey Wolf Optimizer-Based Feature Selection Methods for Classification. In S. Mirjalili, H. Faris, & I. Aljarah (Eds.), Evolutionary Machine Learning Techniques: Algorithms and Applications (pp. 273–286). Springer. https://doi.org/10.1007/978-981-32-9990-0_13

[15] Adamu, A., Abdullahi, M., Junaidu, S. B., & Hassan, I. H. (2021). An hybrid particle swarm optimization with crow search algorithm for feature selection. Machine Learning with Applications, 6, 100108. https://doi.org/10.1016/j.mlwa.2021.100108

[16] Whale optimization algorithm: A systematic review of contemporary applications, modifications and developments | Neural Computing and Applications. (n.d.). Retrieved May 6, 2024, from https://link.springer.com/article/10.1007/s00521-020-04849-z

[17] Ramya, J., Somasundareswari, D., & Vijayalakshmi, P. (2020). Gas chimney and hydrocarbon detection using combined BBO and artificial neural network with hybrid seismic attributes. Soft Computing, 24(3), 2341–2354. https://doi.org/10.1007/s00500-019-04064-6

[18] Ewees, A. A., El Aziz, M. A., & Hassanien, A. E. (2019). Chaotic multi-verse optimizer-based feature selection. Neural Computing and Applications, 31(4), 991–1006. https://doi.org/10.1007/s00521-017-3131-4

[19] Pavithra Guru, R., & Vaithianathan, V. (2020). An efficient VLSI circuit partitioning algorithm based on satin bowerbird optimization (SBO). Journal of Computational Electronics, 19(3), 1232–1248. https://doi.org/10.1007/s10825-020-01491-9

[20] Gokulnath, C. B., & Shantharajah, S. P. (2019). An optimized feature selection based on genetic approach and support vector machine for heart disease. Cluster Computing, 22(6), 14777–14787. https://doi.org/10.1007/s10586-018-2416-4

[21] Jia, W., Sun, M., Lian, J., & Hou, S. (2022). Feature dimensionality reduction: A review. Complex & Intelligent Systems, 8(3), 2663–2693. https://doi.org/10.1007/s40747-021-00637-x

[22] Ahmad, M. W., Reynolds, J., & Rezgui, Y. (2018). Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. Journal of Cleaner Production, 203, 810–821. https://doi.org/10.1016/j.jclepro.2018.08.207

[23] Ekanayake, I. U., Palitha, S., Gamage, S., Meddage, D. P. P., Wijesooriya, K., & Mohotti, D. (2023). Predicting adhesion strength of micropatterned surfaces using gradient boosting models and explainable artificial intelligence visualizations. Materials Today Communications, 36, 106545. https://doi.org/10.1016/j.mtcomm.2023.106545

[24] Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: An interdisciplinary review. Journal of Big Data, 7(1), 94. https://doi.org/10.1186/s40537-020-00369-8

[25] Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geology Reviews, 71, 804–818. https://doi.org/10.1016/j.oregeorev.2015.01.001

[26] Sumayli, A. (2023). Development of advanced machine learning models for optimization of methyl ester biofuel production from papaya oil: Gaussian process regression (GPR), multilayer perceptron (MLP), and K-nearest neighbor (KNN) regression models. Arabian Journal of Chemistry, 16(7), 104833. https://doi.org/10.1016/j.arabjc.2023.104833

[27] Lian, Z., Ma, Y., Li, M., Lu, W., & Zhou, W. (2024). Discovery Precision: An effective metric for evaluating performance of machine learning model for explorative materials discovery. *Computational Materials Science*, *233*, 112738. https://doi.org/10.1016/j.commatsci.2023.112738

Doi : https://doi.org/10.21608/jaiep.2024.354005
Received: March 19, 2024 Accepted: April 22, 2024