Journal of Computer Engineering, Electronics and Information Technology (COELITE)
Journal homepage: https://ejournal.upi.edu/index.php/COELITE

Optimizing Logistic Regression for Digital Marketing Campaigns: Insights from Hyperparameter Tuning with Optuna

Budiman[1], Nur Alamsyah[2], Elia Setiana[3], Fikri Muhamad Fahmi[4]*
[1,3]Informatika, Universitas Informatika dan Bisnis Indonesia, Indonesia
[2,4]Sistem Informasi, Universitas Informatika dan Bisnis Indonesia, Indonesia
Correspondence E-mail: budiman@unibi.ac.id

A B S T R A C T

The shift from mass media advertising to digital marketing driven by the internet revolution highlights the need for data-driven strategies and emerging technologies like Artificial Intelligence (AI). This study aims to develop effective strategies for digital marketing campaigns that enhance customer engagement and increase conversion rates. Key factors such as CampaignChannel, CampaignType, AdSpend, and demographic characteristics (Age, Gender, Income) were analyzed about performance metrics like ConversionRate and WebsiteVisits. A dataset with 20 customer demographics and campaign details features was processed and evaluated using machine learning models, including Logistic Regression, Random Forest, and XGBoost. Pre-processing involved handling missing values, feature selection, and splitting data into training and testing sets. Hyperparameter tuning using Optuna optimized the Logistic Regression Model, achieving the best performance with 89% accuracy. The findings reveal significant relationships between campaign factors and customer behaviour, providing actionable insights to enhance ROI (ROI). This study contributes to a machine learning-based framework for effective segmentation, personalized interactions, and efficient marketing budget management. The study advances AI applications in digital marketing by addressing challenges like data dynamics and shifting business conditions, paving the way for adaptive and data-driven strategies.

A R T I C L E I N F O

Article History:
Submitted/Received 01 Mar 2025
First Revised 15 Mar 2025
Accepted 25 Mar 2025
First Available online 01 Apr 2025
Publication Date 01 Apr 2025

Keyword:
Digital Marketing,
Customer Engagement,
Machine Learning,
Logistic Regression,
Campaign Optimization,
Hyperparameter,
Optuna

1. INTRODUCTION

The Internet revolution and digital marketing have shifted the dominance of mass media advertising towards more modern promotions through the internet and electronic devices in the digital environment [1][2]. In the ever-evolving social network advertising landscape, accurate and quality data play a critical role in the success of predictive models [3]. Businesses must understand customer behaviour and preferences to increase revenue and create better customer engagement [4]. In addition, artificial intelligence (AI) as an emerging technology has become a significant force in various fields, such as healthcare, security, industry, big data, the modern economy, and digital marketing [5]. Studies show that AI will increasingly influence marketing operations by helping companies increase customer loyalty through more personalized interactions [6][7].

AI-powered digital marketing is revolutionizing how organizations create campaign content, generate leads, reduce customer acquisition costs, and manage customer experience efficiently [8]. Algorithms used by online promotion platforms can identify the best strategies to improve campaign performance, although some companies prefer to build customized systems internally [9]. Challenges in digital marketing include dynamic user data and changing business conditions that can affect campaign success [10][11]. Therefore, companies must be able to conduct in-depth customer segmentation by discovering their behavioural patterns to market products more effectively [12].

The integration of artificial intelligence (AI) and machine learning (ML) in digital marketing brings significant changes in traditional marketing strategies [13]. AI helps companies predict consumer demand, optimize advertising campaigns, and find the optimal target audience to increase marketing effectiveness [14][15]. Deep learning through AI allows marketers to tap into user behaviour patterns on websites or mobile apps to respond quickly and refine advertising offers [16][17]. By utilizing these technologies, digital marketing becomes necessary for companies to keep up with evolving consumer needs and opens up new domains for the future of marketing strategies [18].

This study aims to develop effective strategies for digital marketing campaigns that enhance customer engagement and increase conversion rates. This study analyses the influence of factors such as campaign channel, campaign type, ad spend, and customer demographic characteristics (age, gender, and income) on key performance metrics such as conversion rate and website visits. In addition, this study aims to implement and evaluate machine learning models to predict customer behaviour patterns and provide strategic insights that can be applied to improve campaign efficiency and return on investment (ROI).

As a contribution, this study is expected to provide a machine learning-based framework that improves the accuracy of customer behaviour prediction and enables companies to optimize campaign strategies through more effective segmentation, personalization of interactions, and more efficient management of marketing budgets. As such, this study contributes to the scientific literature on digital marketing and the application of artificial intelligence technology in improving the effectiveness of marketing campaigns.

2. METHODS

The methodology depicted in **Figure 1** outlines the sequential steps for optimizing digital marketing campaigns. It starts with a comprehensive review of relevant literature to explore current strategies and identify areas for improvement.

DOI: https://doi.org/10.17509/coelite.v4i1.81020
p- ISSN 2829-4157 e- ISSN 2829-4149

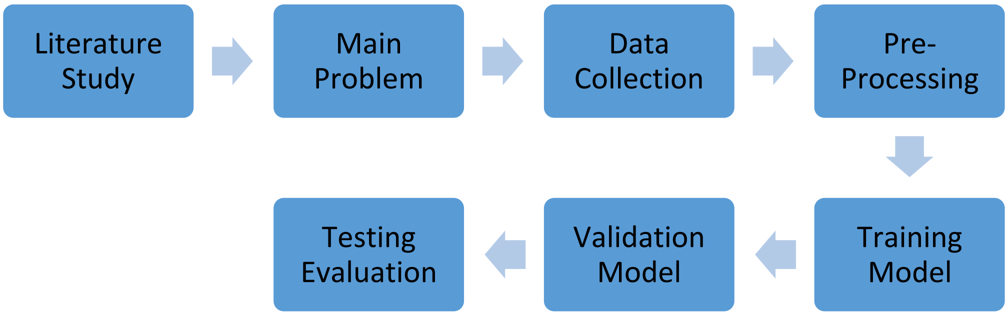

Figure 1. Proposed Method

2.1. Literature Study

This study draws on relevant literature on digital marketing campaign optimization. Previous studies show that metrics such as ClickThroughRate (CTR), ConversionRate, WebsiteVisits, and AdSpend are crucial in understanding customer behaviour and campaign effectiveness. Studies have also highlighted the importance of segmenting campaigns based on demographic features such as Age, Gender, and Income. Furthermore, tools like predictive modelling and machine learning have been widely used to identify patterns in customer interactions and optimize campaign performance [19][20].

2.2. Main Problem

The main problem to be addressed is developing a strategy that maximizes customer engagement and Conversion in digital marketing campaigns. The objective is to analyze the impact of factors such as CampaignChannel, CampaignType, AdSpend, and demographic characteristics (Age, Gender, Income) on key performance metrics like ConversionRate and WebsiteVisits. This study aims to provide actionable insights to enhance campaign efficiency and ROI (Return on Investment).

2.3. Data Collection

We utilized a digital marketing campaign dataset from Kaggle, consisting of 20 features capturing customer demographics, campaign details, and performance metrics. The dataset contains 8,000 records, providing a comprehensive view of customer interactions and campaign performance. The dataset includes the following key features in **Table 1**:

Table 1. Features of a Digital Marketing Campaign

Feature	Description
CustomerID	Unique identifier for each customer.
Age	Age of the customer.
Gender	Gender of the customer (Male/Female).
Income	Income level of the customer.
CampaignChannel	The channel used for the campaign (Social Media, Email, PPC).
CampaignType	Type of campaign (Awareness, Retention, Conversion).
AdSpend	Advertising expenditure in the campaign.
ClickThroughRate	The proportion of ad clicks relative to total impressions.
ConversionRate	The ratio of successful conversions to total interactions.

DOI: https://doi.org/10.17509/coelite.v4i1.81020
p- ISSN 2829-4157 e- ISSN 2829-4149

WebsiteVisits	Number of visits to the website.
PagesPerVisit	Average number of pages viewed per visit.
TimeOnSite	Time spent on the website during visits.
SocialShares	Number of shares on social media.
EmailOpens	Count of emails opened by customers.
EmailClicks	Count of clicks on links within emails.
PreviousPurchases	Total purchases made by the customer before the campaign.
LoyaltyPoints	Customer loyalty points accumulated.
AdvertisingPlatform	Platform used for advertisements.
AdvertisingTool	Specific tools or technologies used in the campaign.
Conversion	Binary indicator for conversion success (1 for success, 0 for failure).

2.4. Pre-Processing

Pre-processing of the dataset involves several steps to ensure data quality and prepare it for analysis:

a. Handling missing values to ensure completeness.
b. Normalizing numerical features like AdSpend, Income, and TimeOnSite to ensure uniformity.
c. Encoding categorical variables such as CampaignChannel and Gender for machine learning compatibility.
d. Feature selection is used to identify the most relevant predictors for Conversion and other key metrics.
e. Splitting the dataset into training (80%) and testing (20%) sets to allow independent Model evaluation.

2.5. Training Model

We trained machine learning models using the pre-processed dataset to predict customer conversions and optimize campaign strategies. Models such as Logistic Regression, Random Forest, XGBoost, and Support Vector Machine were employed to analyze the relationship between demographic and campaign features and key performance indicators. Hyperparameter tuning and cross-validation were applied to enhance model accuracy and generalizability.

2.6. Validation Model

Cross-validation techniques such as k-fold cross-validation were employed to evaluate the robustness of the predictive models. This method helps prevent overfitting by ensuring that each data subset is used for training and testing. The averaged results across folds provide a reliable estimate of model performance on unseen data.

2.7. Testing Evaluation

The final step involved evaluating model performance using accuracy, precision, recall, and F1-score metrics. These metrics provided a comprehensive view of the model's ability to predict conversions and optimize campaign outcomes. Specific focus was given to precision (to minimize false positives in conversion prediction) and recall (to ensure all potential

DOI: https://doi.org/10.17509/coelite.v4i1.81020
p- ISSN 2829-4157 e- ISSN 2829-4149

conversions are identified). The evaluation results inform recommendations for campaign improvement and customer engagement strategies.

3. RESULTS AND DISCUSSION

3.1. Pre-Processing

Table 2 provides a comprehensive overview of the Digital Marketing Campaign Dataset, capturing key customer attributes and performance metrics. The dataset includes details such as Age, which ranges from 32 to 69 years, and Gender, where Male and Female customers are represented, reflecting a diverse target audience. CampaignChannel shows the distribution of marketing strategies across Social Media, Email, and PPC, highlighting the varied platforms utilized for customer engagement.

AdSpend demonstrates significant variation, ranging from $539.53 to $6497.87, indicating differences in budget allocation across campaigns. Metrics like ClickThroughRate and ConversionRate further illustrate campaign effectiveness, with ClickThroughRate varying from 0.0439 to 0.2774 and ConversionRate ranging from 0.0764 to 0.1827. These values suggest diverse levels of customer interaction and conversion success. Additionally, customers with higher AdSpend tend to exhibit higher engagement rates, as observed in Social Media and PPC campaigns.

Engagement indicators such as WebsiteVisits, PagesPerVisit, and TimeOnSite provide insights into customer behaviour on digital platforms. For example, customers with higher TimeOnSite often view more PagesPerVisit, indicating a deeper exploration of campaign content. While some customers exhibit high levels of interaction, others display minimal engagement. SocialShares highlights the extent of content dissemination on social media, varying widely among campaigns, with specific customers sharing content up to 89 times.

Email performance is measured through EmailOpens and EmailClicks, revealing varying levels of customer responsiveness to email marketing. Notably, campaigns with higher EmailClicks generally correlate with better ConversionRate, emphasizing the importance of effective email Communication strategies. LoyaltyPoints, ranging from 688 to 4345, underscore differences in customer loyalty and past interactions, suggesting that loyal customers are more likely to engage with campaigns and convert successfully.

Lastly, the Conversion column identifies whether a customer was successfully converted (1) or not (0), with the majority achieving conversions, reflecting the overall effectiveness of these campaigns. This dataset forms the basis for analyzing and optimizing digital marketing strategies to enhance customer engagement and campaign outcomes. By identifying patterns within this data in **Table 2**, businesses can tailor campaigns to meet customer preferences better and maximize return on investment (ROI).

Table 2. Digital Marketing Campaign Dataset

index	Age	Gender	CampaignChannel	AdSpend	ClickThroughRate	ConversionRate
0	56	Female	Social Media	6497.87	0.0439	0.0880
1	69	Male	Email	3898.67	0.1557	0.1827
2	46	Female	PPC	1546.43	0.2774	0.0764
3	32	Female	PPC	539.53	0.1376	0.0880
4	60	Female	PPC	1678.04	0.2529	0.1099
298	45	Male	Social Media	1234.67	0.1673	0.0642
299	38	Male	Email	2345.78	0.1267	0.0493

DOI: https://doi.org/10.17509/coelite.v4i1.81020
p- ISSN 2829-4157 e- ISSN 2829-4149

index	Age	Gender	CampaignChannel	AdSpend	ClickThroughRate	ConversionRate
300	32	Female	PPC	3467.98	0.2100	0.0810
301	58	Female	Social Media	5467.43	0.1900	0.1023
302	61	Male	Email	2845.65	0.1450	0.0710

index	WebsiteVisits	PagesPerVisit	TimeOnSite	SocialShares	EmailOpens
0	0	2.40	7.40	19	6
1	42	2.92	5.35	5	2
2	2	8.22	13.79	0	11
3	47	4.54	14.69	89	2
4	0	2.05	13.99	6	6
298	15	1.78	6.23	4	4
299	20	2.89	5.15	8	5
300	10	3.12	7.85	15	6
301	30	4.30	10.45	20	10
302	25	3.80	8.40	12	4

index	EmailClicks	LoyaltyPoints	Conversion
0	9	688	1
1	7	3459	1
2	2	2337	1
3	2	2463	1
4	6	4345	1
298	2	1234	0
299	3	2100	0
300	4	2340	0
301	7	3000	0
302	5	2500	0

Table 3 shows the descriptive statistics of the dataset that provide deep insights into customer attributes and digital marketing campaign performance. Regarding demographics, the average age of customers is 43.6 years old, with an age range from 18 to 69 years old, reflecting a wide range of targets. Customer income varied significantly, with an average of $84,664.20, but spread from $20,014 to $149,986, suggesting segmentation by purchasing power. Average advertising expenditure was recorded at $5,000.94 with significant variations (standard deviation of $2,838.04), reflecting different budget allocations between campaigns. The average ClickThroughRate (CTR) was 0.1548, and the average ConversionRate was 0.1044, highlighting the moderate effectiveness of campaigns in attracting and converting customers. These metrics show that campaigns with a higher CTR tend to generate a higher ConversionRate, although there is some variation.

Customer engagement is evident from the average website visit (24.75 visits) and average time spent (7.73 minutes). Customers who spend more time on the site tend to visit more pages per visit (5.55 pages on average). Social media sharing activity was also significant, with an average of 49.8 shares per campaign, demonstrating the appeal of the content on social media platforms. Customer response to email campaigns was also good, with an average of

DOI: https://doi.org/10.17509/coelite.v4i1.81020
p- ISSN 2829-4157 e- ISSN 2829-4149

9.48 email opens and 4.47 clicks on email links. Customers with more email engagement tend to have a higher likelihood of Conversion, as seen from the positive correlation between EmailClicks and ConversionRate. Customer loyalty was also an essential factor, with an average of 2,490.27 loyalty points, indicating an ongoing relationship between the customer and the brand. The overall campaign conversion rate was very high, with an average of 87.65% of subscribers successfully converted. This reflects excellent success in the overall digital campaign strategy. This analysis shows that optimizing ad spend, email marketing, and website engagement can improve campaign performance to maximize ROI and strengthen customer relationships.

Table 3. Descriptive Statistics of the Digital Marketing Campaign Dataset

Feature	mean	std	min	25%	50%	75%	max
Age	43.63	14.90	18	31	43	56	69
Income	84664.20	37580.39	20014	51744.5	84926.5	116815.75	149986
AdSpend	5000.94	2838.04	100.05	2523.22	5013.44	7407.99	9997.91
ClickThroughRate	0.15	0.08	0.01	0.08	0.15	0.23	0.30
ConversionRate	0.10	0.05	0.01	0.06	0.10	0.15	0.20
WebsiteVisits	24.75	14.31	0	13	25	37	49
PagesPerVisit	5.55	2.61	1.00	3.30	5.53	7.84	10.00
TimeOnSite	7.73	4.23	0.50	4.07	7.68	11.48	15.00
SocialShares	49.80	28.90	0	25	50	75	99
EmailOpens	9.48	5.71	0	5	9	14	19
EmailClicks	4.47	2.86	0	2	4	7	9
PreviousPurchases	4.49	2.89	0	2	4	7	9
LoyaltyPoints	2490.27	1429.53	0	1254.75	2497	3702.25	4999
Conversion	0.88	0.33	0	1	1	1	1

Figure 2 is a heatmap image that visualizes the correlation relationship between the various variables in the digital marketing campaign dataset. The correlation values range from -1 to 1, with 1 indicating a perfect positive correlation, -1 indicating a perfect negative correlation, and 0 indicating no relationship. Some important patterns can be observed from this heatmap. The ConversionRate variable shows a moderate positive relationship with ClickThroughRate, which suggests that the more customers click on an advert, the more likely they are to convert. In addition, PagesPerVisit and TimeOnSite have a positive correlation, indicating that customers who spend more time on the website tend to browse more pages, which can increase the chances of engagement.

AdSpend has a low correlation with ConversionRate, which may indicate that ad spend alone is insufficient to ensure a high conversion rate. This underlines the importance of other factors, such as the quality of ad content and audience segmentation strategies. The LoyaltyPoints variable has a moderate positive correlation with Conversion, suggesting that more loyal customers tend to convert more quickly. Some variables, such as Income and SocialShares, show low or almost zero correlation with key performance variables such as

DOI: https://doi.org/10.17509/coelite.v4i1.81020
p- ISSN 2829-4157 e- ISSN 2829-4149

ConversionRate and ClickThroughRate, indicating that customer income and social media sharing activity may not directly affect campaign performance in this dataset. Overall, this heatmap shows that while some variables such as ClickThroughRate, PagesPerVisit, and LoyaltyPoints have a significant influence on campaign performance, many other variables show weaker relationships, highlighting the need for additional analysis to identify key factors that influence marketing campaign success.

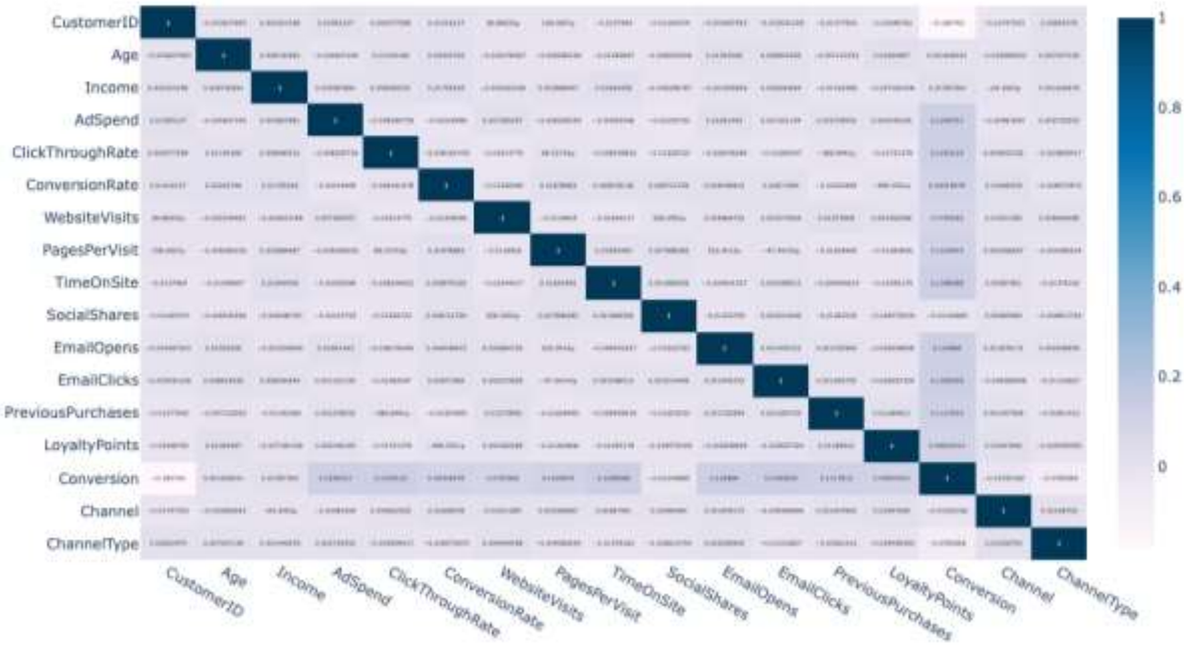

Figure 2. Heatmap

Figure 3 above shows the results of the Chi-Square analysis to measure the strength of the relationship between categorical variables in the dataset. Higher Chi-Square values indicate a stronger relationship between that variable and the target or dependent variable in the analysis. In this graph, Gender, CampaignChannel, and Age have the highest Chi-Square values, indicating that these variables significantly influence the target or campaign outcome. This suggests that demographic factors such as gender and age, as well as the campaign channel used, play an essential role in determining the success of a digital marketing campaign. In contrast, variables such as SocialShares, ChannelType, and PreviousPurchases have lower Chi-Square values, suggesting that the relationship between these variables and campaign outcomes is relatively weak or insignificant. Variables such as AdvertisingPlatform and AdvertisingTool showed almost no relationship, indicating that they may not be relevant or have little impact on campaign success. These results provide valuable insights for identifying essential variables that should be prioritized in marketing strategies, such as focusing on customer demographics and selecting appropriate campaign channels while reducing attention to variables with weak influence for better efficiency.

DOI: https://doi.org/10.17509/coelite.v4i1.81020
p- ISSN 2829-4157 e- ISSN 2829-4149

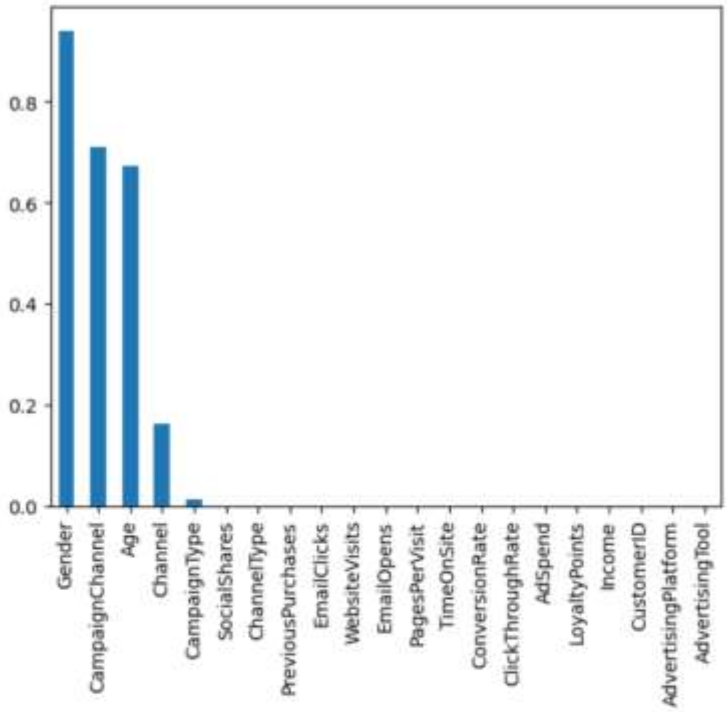

Figure 3. Chi-Square analysis

3.2. Training Model

Table 4 compares the classification reports; the performance of the tested models shows significant variation in their ability to handle predictions. Support Vector Classification (SVC) showed great weakness in predicting class 0, with precision, recall, and f1-score being 0.00 for this class. This indicates that the model could not detect or predict class 0 despite performing very well for class 1 (precision: 0.88, recall: 1.00, f1-score: 0.93). The model's overall accuracy is 0.88, but the imbalance in class prediction makes it less reliable in applications that require detection in both classes.

In contrast, Random Forest Classification showed better ability in detecting class 0 with a precision of 0.97, although its recall was low at 0.14, resulting in an f1-score of 0.25. For class 1, the model performed almost flawlessly with a recall of 1.00 and an f1-score of 0.94, resulting in an overall accuracy of 0.89. While the accuracy and performance of class 1 are pretty high, the challenge remains in improving the recall of class 0 to provide more balanced results. While both models show excellence in handling class 1 (majority), Random Forest Classification has the edge in accommodating class 0 predictions, although there is still room for improvement. Support Vector Classification, however, requires further tuning to handle significant class imbalance. In real applications, model selection depends on business priorities, such as whether it is more critical to maximize minority class predictions or maintain high accuracy on the majority class.

Table 4. Comparison of Classification Report

Metric	LR	SVM	RF	KNN	DT
Precision (Class 0)	0.77	0	0.97	0.52	0.54
Recall (Class 0)	0.17	0	0.14	0.17	0.56

DOI: https://doi.org/10.17509/coelite.v4i1.81020
p- ISSN 2829-4157 e- ISSN 2829-4149

Metric	LR	SVM	RF	KNN	DT
F1-Score (Class 0)	0.28	0	0.25	0.25	0.55
Precision (Class 1)	0.9	0.88	0.89	0.89	0.94
Recall (Class 1)	0.99	1	1	0.98	0.93
F1-Score (Class 1)	0.94	0.93	0.94	0.93	0.93
Accuracy	0.89	0.88	0.89	0.88	0.89
Macro Avg Precision	0.83	0.44	0.93	0.71	0.74
Macro Avg Recall	0.58	0.5	0.57	0.57	0.75
Macro Avg F1-Score	0.61	0.47	0.6	0.59	0.74
Weighted Avg F1-Score	0.86	0.82	0.86	0.85	0.89

The learning curve for Logistic Regression shows a steady increase in validation accuracy as the dataset size increases, reaching a maximum accuracy of around 89%. Training accuracy is slightly higher than validation accuracy, but the gap between the two is relatively small, indicating that the model has low bias and only slight overfitting. The model shows consistent performance, making it a good choice for datasets with fairly linear or straightforward patterns.

The Support Vector Classifier (SVC) has a more volatile learning curve. The training accuracy remains high, but the validation accuracy is relatively lower and does not show significant improvement as the data grows. The large gap between training and validation accuracy indicates serious overfitting. This suggests that the SVC is less than optimal in generalization on this dataset, mainly if no further parameter tuning is performed.

Random Forest shows almost perfect training accuracy across all dataset sizes, but validation accuracy tends to plateau at around 89%. The large gap between training and validation accuracy indicates that the model suffers from significant overfitting. Despite the high validation accuracy, the model may be less efficient because it learns too many details from the training data, which may affect the generalization of new data.

The K-Nearest Neighbor (KNN) learning curve shows a promising upward trend in validation accuracy as the data grows, approaching 91% at the maximum dataset size. Training accuracy is slightly higher than validation, but the gap between the two is relatively small, indicating that the model has a good balance between bias and variance. KNN appears to be one of the best models for generalization on this dataset.

Decision Tree has a similar learning curve to Random Forest, with training accuracy close to 100%, but validation accuracy tends to plateau at around 89%. The large gap between training and validation accuracy suggests that the Decision Tree also suffers from overfitting, even though it is a simpler model than Random Forest. This indicates that Decision Tree may be less effective for this dataset without additional regularization methods.

Based on the learning curve analysis in **Figure 4**, K-Nearest Neighbor (KNN) and Logistic Regression performed best in generalization, with a small gap between training and validation accuracy and high validation accuracy. Random Forest and Decision Tree have overfitting issues, although their validation accuracy is quite good. Meanwhile, SVC performs suboptimal due to low validation accuracy and significant fluctuations. Since Logistic Regression has better accuracy than SVC at 0.89, Logistic Regression is the best choice for this dataset, depending on the complexity of the data and the computational resources available.

DOI: https://doi.org/10.17509/coelite.v4i1.81020
p- ISSN 2829-4157 e- ISSN 2829-4149

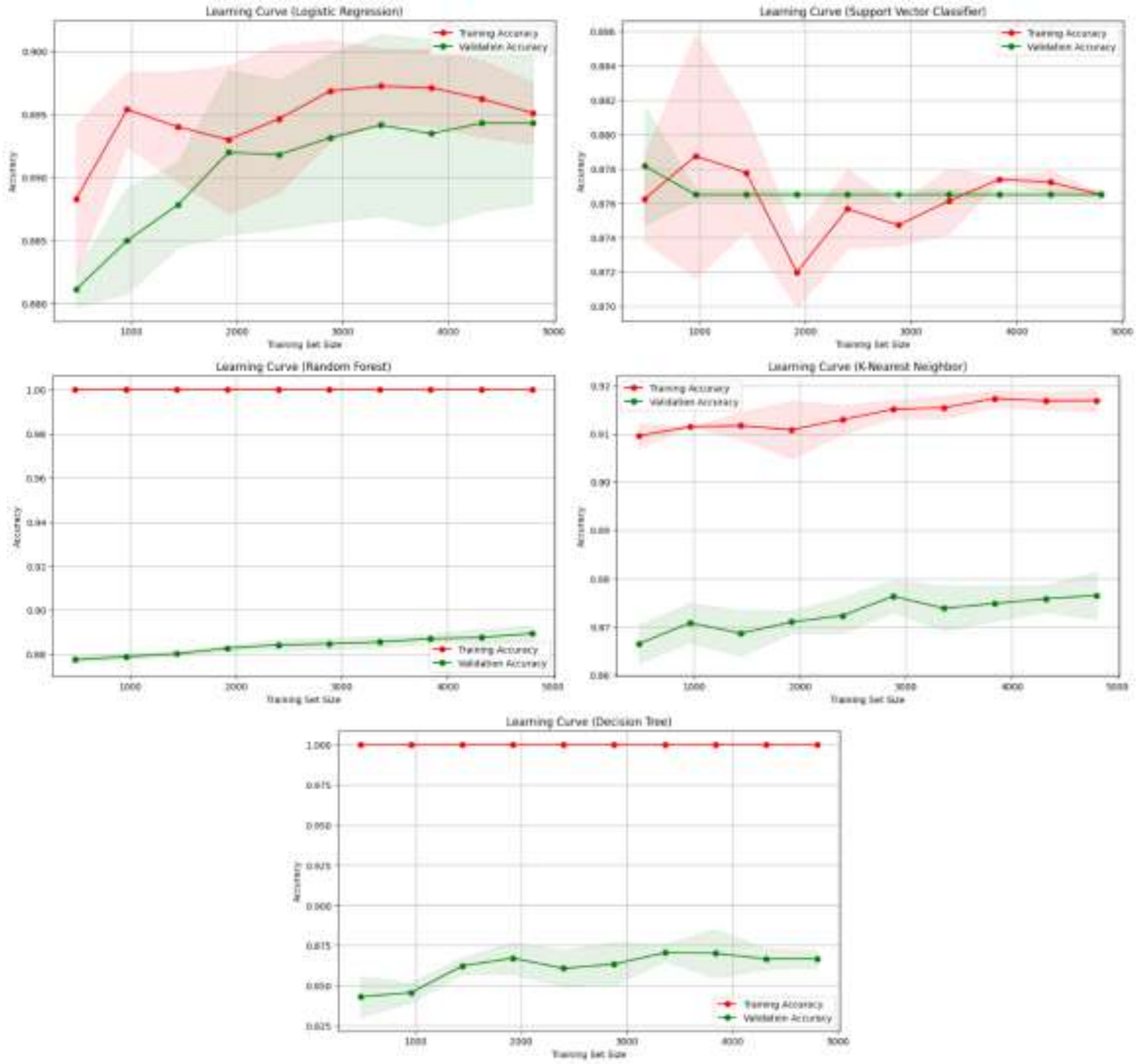

Figure 4. Learning Curve

After the analysis shows that Logistic Regression is the best model based on the learning curve, the next step is to improve the performance of this model through hyperparameter tuning. Hyperparameter tuning aims to find the optimal combination of parameters so that the model can maximize accuracy and generalization. **Figure 5** shows the learning curve analysis of the hyperparameter tuning process of the Logistic Regression model using Grid Search CV, Bayes Search CV, and Optuna, showing significant differences in model performance concerning training and validation accuracy. The first graph with Grid Search CV shows a stable trend with training and validation accuracy slowly approaching equilibrium as the dataset size increases. However, the gap between training and validation accuracy is still visible at small dataset sizes, indicating the model has a slight overfitting that decreases with larger datasets.

The learning curve of Bayes Search CV shows better performance than Grid Search, with more stable and higher validation accuracy. The training and validation accuracies on larger datasets tend to converge, indicating that Bayesian Optimisation successfully finds better parameter combinations to improve model generalization. In addition, the fluctuation in validation accuracy is smaller than that of Grid Search, reflecting more consistent results in tuning iterations.

Optuna, as a more adaptive hyperparameter tuning approach, showed very competitive results. The validation accuracy is higher than the other methods on almost all dataset sizes, with a small gap between training and validation accuracy. This shows that the Logistic Regression model tuned using Optuna successfully minimizes overfitting while maintaining high validation accuracy. Optuna's adaptivity in efficiently exploring the parameter space provides a significant advantage over Grid Search and Bayes Search CV.

Optuna produced the most optimal model for Logistic Regression in generalisability and stability, followed by Bayes Search CV. Grid Search CV, although providing decent results, is less efficient and performs less than the other two methods. Thus, Optuna is the recommended method for hyperparameter tuning in Logistic Regression based on this dataset and learning curve analysis.

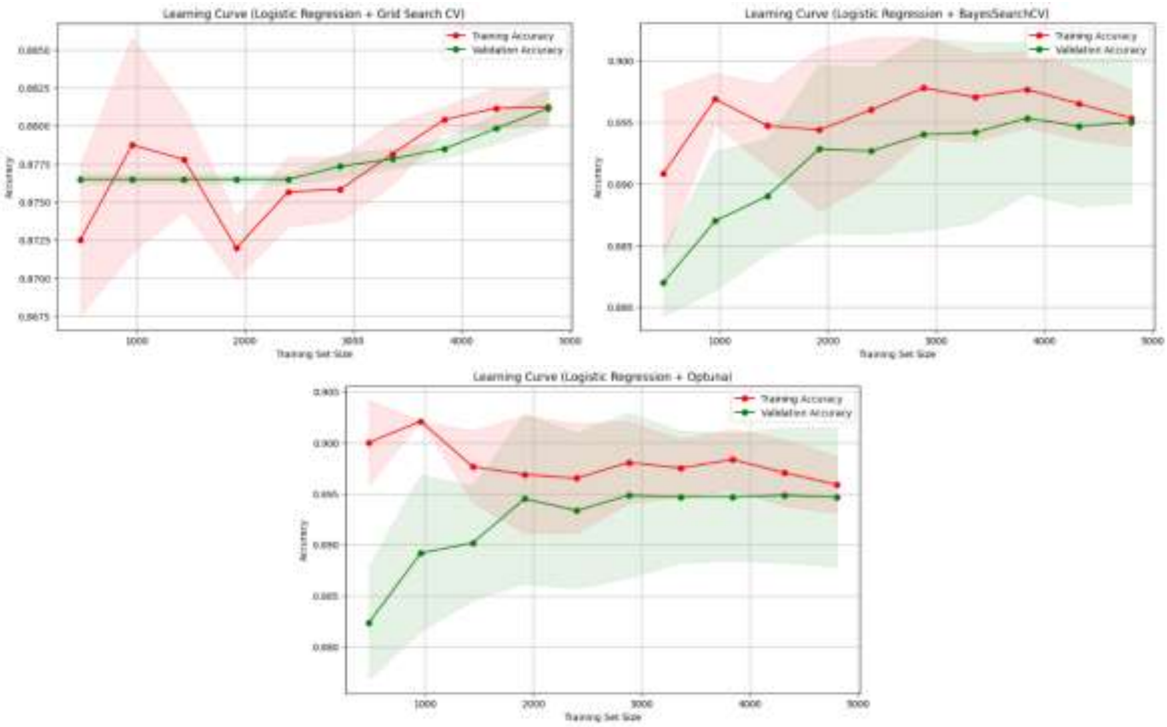

Figure 5. Learning Curve Analysis of the Hyperparameter Tuning Process of the Logistic Regression

After hyperparameter tuning using Optuna on Logistic Regression, model performance was slightly improved, especially for Class 0. Precision for Class 0 decreased from 0.77 to 0.75, but recall increased from 0.17 to 0.18, which resulted in an increase in F1-Score from 0.28 to 0.29. This shows that although the model slightly sacrificed precision in predicting Class 0, the model's ability to detect more Class 0 samples increased. For Class 1, the precision, recall, and F1-Score remained consistent at 0.90, 0.99, and 0.94, respectively, indicating that the model's performance for Class 1 was unaffected by tuning. The model accuracy remained at 89%, with no significant change. The best parameters found by Optuna are {'penalty': 'l1', 'C': 294.9937556473663, 'solver': 'saga'}, which provides the optimal configuration for the Logistic Regression model.

Regarding macro averages, recall increased from 0.58 to 0.59, indicating a slight improvement in the model's ability to detect samples on average in both classes. The F1-Score macro average also saw a slight increase from 0.61 to 0.62, reflecting a small improvement in the balance of precision and recall. However, the weighted average of precision, recall, and F1-Score remained stable at 0.88, 0.89, and 0.86, respectively, indicating that the contribution of Class 1 dominating the final model results was still very strong.

DOI: https://doi.org/10.17509/coelite.v4i1.81020
p- ISSN 2829-4157 e- ISSN 2829-4149

Compared to Logistic Regression before tuning, hyperparameter tuning with Optuna slightly improved, especially in detecting samples in Class 0, which was previously a significant weakness of the model. However, this improvement was insignificant and did not affect the overall accuracy. In conclusion, although hyperparameter tuning provides a slight improvement, additional strategies, such as handling data imbalance by oversampling or undersampling, are needed to improve performance on Class 0 more substantially.

3.3. Validation Model

Table 5 above is cross-validation results show fairly consistent performance across folds, with metrics such as precision, recall, and f1-score showing good stability. The model performed exceptionally well in the majority class (Class 1), with an average accuracy of 0.89 and an average recall of 0.99. This shows the model can minimize false positives while detecting almost all Class 1 samples well. The average F1-score for Class 1, which reached 0.94, reflects the optimal balance between precision and recall. In contrast, the performance of the minority class (Class 0) was much lower, with an average recall of only 0.20, indicating that the model could only detect about 20% of the total Class 0 samples. Precision for Class 0 also averaged 0.77, suggesting that predictions for Class 0 still generated many false positives.

The model's overall accuracy averages 89.3%, indicating that most samples can be classified correctly. However, this accuracy is heavily influenced by the dominant performance of Class 1, which has a much larger data distribution than Class 0. This imbalance means that high accuracy does not necessarily reflect good performance for both classes equally. Analysis of the mean and standard deviation shows that the performance for Class 1 is very stable, with almost insignificant variation between folds. In contrast, the recall for Class 0 has a higher standard deviation (0.0126), indicating that the model performance for this class is more sensitive to data variations.

Table 5. Classification Metrics by Fold

Metric	Fold 0	Fold 1	Fold 2	Fold 3	Mean	Std. Dev
tn	45	52	51	53		
fp	202	195	196	194		
fn	6	18	15	26		
tp	1747	1735	1738	1727		
prec_1	0.90	0.90	0.90	0.90	0.90	0.0011
prec_0	0.88	0.74	0.77	0.67	0.78	0.076092
recall_1	0.99	0.99	0.99	0.99	0.99	0.004081
recall_0	0.18	0.21	0.21	0.21	0.20	0.012601
accuracy	0.90	0.89	0.89	0.89	0.89	0.002208
f1_score	0.94	0.94	0.94	0.94	0.94	0.001347

In conclusion, while the model performs very well in Class 1, its performance in Class 0 requires further attention. To improve the performance of this minority class, measures such as handling data imbalance using oversampling (such as SMOTE) or undersampling, as well as adjusting the class weights in the loss function, can be applied. In addition, hyperparameter tuning that focuses more on improving recall for Class 0 can help to correct this performance imbalance. Using these strategies, the model is expected to achieve a better balance between the two classes, resulting in a more optimized classification.

DOI: https://doi.org/10.17509/coelite.v4i1.81020
p- ISSN 2829-4157 e- ISSN 2829-4149

3.4. Testing Evaluation

The final step involved evaluating model performance using accuracy, precision, recall, and F1-score metrics. These metrics provide a comprehensive view of the model's ability to predict conversions and optimize campaign outcomes. Specific focus was given to precision, which minimizes false positives in conversion prediction, and recall, which ensures that all potential conversions are identified. The model exhibited slight improvements after hyperparameter tuning with Optuna, particularly for Class 0. Precision for Class 0 decreased slightly from 0.77 to 0.75, but recall improved from 0.17 to 0.18, leading to an F1-score increase from 0.28 to 0.29. For Class 1, the precision, recall, and F1-score remained stable at 0.90, 0.99, and 0.94, respectively. This stability indicates that the tuning did not negatively affect the performance of the majority class while enhancing the detection of the minority class.

The learning curve analysis from the hyperparameter tuning process showed that Optuna consistently outperformed other methods, such as Grid Search CV and Bayes Search CV, by achieving higher validation accuracy and minimizing overfitting. Optuna successfully balanced training and validation performance, making it the most effective tuning method for Logistic Regression in this context. Despite these improvements, overall accuracy remained at 89%, highlighting the importance of further strategies to address performance in minority classes. Cross-validation results reaffirmed the model's stability, with minimal variations across folds for Class 1 metrics but a higher standard deviation for recall on Class 0, indicating sensitivity to data variations.

These results suggest that while the model performs exceptionally well for Class 1, additional measures are needed to improve performance in Class 0. Techniques such as handling data imbalance through oversampling (e.g., SMOTE) or undersampling and adjusting class weights in the loss function can help address this imbalance. Moreover, targeted hyperparameter tuning focusing on increasing recall for Class 0 could further enhance the model's ability to detect conversions within the minority class. The evaluation insights from this process directly inform recommendations for optimizing campaign strategies and improving customer engagement outcomes. By implementing these strategies, the model is expected to balance both classes better, resulting in a more effective and generalized classification system.

4. CONCLUSION

This study successfully achieved its goal of developing optimal strategies for digital marketing campaigns by increasing customer engagement and conversion rates. The analysis showed that factors such as campaign channel (CampaignChannel), campaign type (CampaignType), advertising spend (AdSpend), as well as customer demographic characteristics (age, gender, and income) have a significant influence on key performance metrics such as ConversionRate and WebsiteVisits. Using a comprehensive digital marketing campaign dataset, this study identifies customer behaviour patterns and generates strategic insights that can be used to improve campaign efficiency and return on investment (ROI).

Machine learning model evaluation results show that Logistic Regression is the best model based on the learning curve and overall performance, with 89% accuracy. Logistic Regression showed a good balance between accuracy, recall, and generalisability. The model was also more stable after hyperparameter tuning using the Optuna method, which resulted in optimal parameter configurations such as {'penalty': 'l1', 'C': 294.9938, 'solver': 'saga'}.

DOI: https://doi.org/10.17509/coelite.v4i1.81020
p- ISSN 2829-4157 e- ISSN 2829-4149

Hyperparameter tuning successfully improved the detection of the minority class (Class 0), although the improvement was still limited.

As a contribution, this study provides a machine learning-based framework that enables the personalization of customer interactions, more effective segmentation, and more efficient marketing budget management. The results of this study not only reinforce the relevance of digital marketing in the modern era but pave the way for the application of artificial intelligence in creating more adaptive and data-driven marketing strategies.

5. ACKNOWLEDGMENT

The authors thank all the individuals and organizations who contributed to this study. Thanks to Universitas Informatika dan Bisnis Indonesia for providing the necessary resources and support throughout the study. The authors are also thankful to the reviewers and peers for their valuable feedback, which greatly enhanced the quality of this work.

6. AUTHORS' NOTE

The authors declare no conflicts of interest related to this study. All data and methodologies used in this study comply with ethical standards and were transparently conducted. The authors take full responsibility for errors or oversights and welcome constructive feedback from the academic and professional community.

7. REFERENCES

[1] B. Peyravi, J. Nekrošienė, and L. Lobanova, 'Revolutionised Technologies for Marketing: Theoretical Review With Focus On Artificial Intelligence', *Verslas: teorija ir praktika*, vol. 21, no. 2, pp. 827–834, 2020.

[2] Q. Yang, P. Li, X. Xu, Z. Ding, W. Zhou, and Y. Nian, 'A Comparative Study on Enhancing Prediction in Social Network Advertisement Through Data Augmentation', in *2024 4t,'International Conference on Machine Learning and Intelligent Systems Engineering (MLISE)*, Jun. 2024, pp. 214–218. doi: 10.1109/MLISE62164.2024.10674203.

[3] 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, doi: 10.1109/TEM.2023.3348991.

[4] B. S. Arasu, B. J. B. Seelan, and N. Thamaraiselvan, 'A Machine Learning-Based Approach To Enhancing Social Media Marketing,' *Computers & Electrical Engineering*, vol. 86, p. 106723, Sep. 2020, doi: 10.1016/j.compeleceng.2020.106723.

[5] M. Boujrad and Y. el M. N. lamlili, 'A New Artificial Intelligence-Based Strategy for Digital Marketing Reinforcement', in *Innovations in Smart Cities Applications Volume 4*, M. Ben Ahmed, İ. Rakıp Karaș, D. Santos, O. Sergeyeva, and A. A. Boudhir, Eds., Cham: Springer International Publishing, 2021, pp. 689–699. doi: 10.1007/978-3-030-66840-2_52.

[6] R. Sarath Kumar Boddu, A. A. Santoki, S. Khurana, P. Vitthal Koli, R. Rai, and A. Agrawal, 'An Analysis To Understand The Role Of Machine Learning, Robotics And Artificial Intelligence In Digital Marketing', *Materials Today: Proceedings*, vol. 56, pp. 2288–2292, Jan. 2022, doi: 10.1016/j.matpr.2021.11.637.

[7] M. Marić, O. Grljević, and L. Gluščević, 'Application Of Artificial Intelligence In Digital Marketing', *Anali Ekon fak Subotica*, no. 52, pp. 21–37, 2024, doi: 10.5937/AnEkSub2300033M.

[8] P. van Esch and J. Stewart Black, 'Artificial Intelligence (AI): Revolutionizing Digital

DOI: https://doi.org/10.17509/coelite.v4i1.81020
p- ISSN 2829-4157 e- ISSN 2829-4149

Marketing', *Australasian Marketing Journal*, vol. 29, no. 3, pp. 199–203, Aug. 2021, doi: 10.1177/18393349211037684.

[9] D. Dumitriu and M. A.-M. Popescu, 'Artificial Intelligence Solutions for Digital Marketing', *Procedia Manufacturing*, vol. 46, pp. 630–636, Jan. 2020, doi: 10.1016/j.promfg.2020.03.090.

[10] N. Saud Abd, O. Salim Atiyah, M. Taher Ahmed, and A. Bakhit, 'Digital Marketing Data Classification by Using Machine Learning Algorithms', *IJEEE*, vol. 20, no. 1, pp. 245–256, Jun. 2024, doi: 10.37917/ijeee.20.1.23.

[11] K. Singhal, J. N. Singh, and V. Sharma, 'Enabling Autonomous Digital Marketing: A Machine Learning Approach for Consumer Demand Forecasting', in *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*, Feb. 2024, pp. 1903–1908. doi: 10.1109/IC2PCT60090.2024.10486327.

[12] C. Wang, 'Efficient Customer Segmentation In Digital Marketing Using Deep Learning With Swarm Intelligence Approach', *Information Processing & Management*, vol. 59, no. 6, p. 103085, Nov. 2022, doi: 10.1016/j.ipm.2022.103085.

[13] V. Singh, B. Nanavati, A. K. Kar, and A. Gupta, 'How to Maximize Clicks for Display Advertisement in Digital Marketing? A Reinforcement Learning Approach', *Inf Syst Front*, vol. 25, no. 4, pp. 1621–1638, Aug. 2023, doi: 10.1007/s10796-022-10314-0.

[14] M. Zeeshan and K. Saxena, 'Explorative Study of Artificial Intelligence in Digital Marketing', in *Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI - 2019)*, A. P. Pandian, R. Palanisamy, and K. Ntalianis, Eds., Cham: Springer International Publishing, 2020, pp. 968–978. doi: 10.1007/978-3-030-43192-1_107.

[15] Q. Han, C. Lucas, E. Aguiar, P. Macedo, and Z. Wu, 'Towards Privacy-Preserving Digital Marketing: An Integrated Framework For User Modeling Using Deep Learning On A Data Monetization Platform', *Electron Commer Res*, vol. 23, no. 3, pp. 1701–1730, Sep. 2023, doi: 10.1007/s10660-023-09713-5.

[16] Y. Lahbabi, S. Raki, H. C. Lamrani, and S. Dehbi, 'Machine Learning In Digital Marketing.', *MENACIS2021*, Nov. 2021, [Online]. Available: https://aisel.aisnet.org/menacis2021/28

[17] M. S. Ullal, I. T. Hawaldar, R. Soni, and M. Nadeem, 'The Role of Machine Learning in Digital Marketing', *Sage Open*, vol. 11, no. 4, p. 21582440211050394, Oct. 2021, doi: 10.1177/21582440211050394.

[18] G. Gürsoy, A. Varol, and S. Varol, 'Impact of Machine Learning in Digital Marketing Applications', in *2022 3rd International Informatics and Software Engineering Conference (IISEC)*, Dec. 2022, pp. 1–6. doi: 10.1109/IISEC56263.2022.9998211.

[19] Imani, M., & Arabnia, H. R. (2023). Hyperparameter Optimization and Combined Data Sampling Techniques In Machine Learning For Customer Churn Prediction: *A Comparative Analysis. Technologies, 11*(6), 167. https://doi.org/10.3390/technologies11060167

[20] Feroz, S. B., Sharmin, N., & Sevas, M. S. (2024). An Empirical Analysis Of Hyperparameter Tuning Impact On Ensemble Machine Learning Algorithm For Earthquake Damage Prediction. Asian Journal of Civil Engineering, 25(4), 3521–3547.

DOI: https://doi.org/10.17509/coelite.v4i1.81020
p- ISSN 2829-4157 e- ISSN 2829-4149