Data and Metadata. 2025; 4:1061
doi: 10.56294/dm20251061

ORIGINAL

Predictive Analytics in Digital Marketing: A Statistical Modeling Approach for Predicting Consumer Behavior

Análisis Predictivo en Marketing Digital: Un Enfoque de Modelación Estadística para la Predicción del Comportamiento del Consumidor

Jazmín Isabel García-Guerra[1], Héctor Oswaldo Aguilar-Cajas[1], Heidy Elizabeth Vergara-Zurita[1], Ana Lucía Rivera-Abarca[1], Freddy Armijos-Arcos[2], José Israel López-Pumalema[3]

[1]Escuela Superior Politécnica de Chimborazo (ESPOCH), Riobamba - Ecuador.
[2]Ingeniería, Diseño y Consultoría (IDYC Cia. Ltda.)
[3]Universidad Estatal Península de Santa Elena.

Cite as: García-Guerra JI, Aguilar-Cajas HO, Vergara-Zurita HE, Rivera-Abarca AL, Armijos-Arcos F, López-Pumalema JI. Predictive Analytics in Digital Marketing: A Statistical Modeling Approach for Predicting Consumer Behavior. Data and Metadata. 2025; 4:1061. https://doi.org/10.56294/dm20251061

Submitted: 03-10-2024 **Revised:** 18-02-2025 **Accepted:** 09-06-2025 **Published:** 10-06-2025

Editor: Dr. Adrián Alejandro Vitón Castillo

Corresponding author: Jazmín Isabel García-Guerra

ABSTRACT

Introduction: the evolution of predictive analytics in digital marketing is deeply rooted in the development of statistical modeling and data analytics.
Objectives: the objective of the present research was to analyze the use of advanced statistical models for predicting consumer behavior in digital marketing environments, highlighting the relevance of predictive analytics in data- driven strategic decision making.
Method: five machine learning, logistic regression, decision tree, random forest, support vector machines (SVM) and neural networks models were evaluated on a synthetic dataset representative of digital consumers belonging to Generation Z. The analysis considered key metrics such as overall accuracy, cross-validation mean and standard deviation, in order to measure both the effectiveness and stability of each model.
Results: the results showed that the logistic regression, Random Forest, SVM and neural network models achieved an accuracy of 97 % with overall consistency (standard deviation of 0,0), positioning them as reliable tools for predicting consumption trends. In contrast, the decision tree showed lower accuracy (92 %) and higher variability, which limits its applicability in complex scenarios.
Conclusion: the study concludes that the combination of accuracy and stability is essential for the implementation of effective predictive models in digital marketing and also highlights the importance of integrating these models into campaign automation and personalization systems to anticipate preferences, improve customer experience and optimize resources.

Keywords: Predictive Analytics; Digital Marketing; Consumer Behavior; Machine Learning; Statistical Modeling.

RESUMEN

Introducción: la evolución del análisis predictivo en el marketing digital está profundamente arraigada en el desarrollo de la modelización estadística y el análisis de datos.
Objetivo: el objetivo de la presente investigación fue analizar el uso de modelos estadísticos avanzados para la predicción del comportamiento del consumidor en entornos de marketing digital, destacando la relevancia del análisis predictivo en la toma de decisiones estratégicas basadas en datos.

Método: se evaluaron cinco modelos de aprendizaje automático, regresión logística, árbol de decisión, bosque aleatorio, máquinas de vectores de soporte (SVM) y redes neuronales sobre un conjunto sintético de datos representativo de consumidores digitales pertenecientes a la Generación Z. El análisis consideró métricas clave como la precisión global (accuracy), la media de validación cruzada y la desviación estándar, con el objetivo de medir tanto la eficacia como la estabilidad de cada modelo.
Resultados: los resultados mostraron que los modelos de regresión logística, Random Forest, SVM y redes neuronales alcanzaron una precisión del 97 % con consistencia total (desviación estándar de 0,0), posicionándolos como herramientas confiables para la predicción de tendencias de consumo. En cambio, el árbol de decisión evidenció una menor precisión (92 %) y mayor variabilidad, lo que limita su aplicabilidad en escenarios complejos.
Conclusión: el estudio concluye que la combinación de precisión y estabilidad es esencial para la implementación de modelos predictivos efectivos en marketing digital también destaca la importancia de integrar estos modelos en sistemas de automatización y personalización de campañas, para anticipar preferencias, mejorar la experiencia del cliente y optimizar recursos.

Palabras clave: Análisis Predictivo; Marketing Digital; Comportamiento del Consumidor; Aprendizaje Automático; Modelado Estadístico.

INTRODUCTION

The evolution of predictive analytics in digital marketing is deeply rooted in the development of statistical modeling and data analytics. At its core, predictive analytics employs historical data to predict future outcomes, leveraging various statistical models, machine learning techniques, and data mining methods to analyze trends and patterns in consumer behavior.[1,2]

The integration of these methodologies has transformed the marketing research landscape, particularly with the advent of Big Data, which has significantly improved the ability to gather and analyze large amounts of consumer information.[3,4]

Historically, the basis of predictive analytics can be traced back to the earliest applications of statistical techniques in marketing. Companies began recognizing the value of data-driven decision-making, and secondary data from reports and market research was emphasized. While less direct than primary data collection, this approach provided marketers with a broader perspective on market trends and consumption patterns.[5,6]

The advent of more sophisticated data processing capabilities allowed marketers to conduct both quantitative and qualitative analyses. Quantitative analysis focuses on numerical data to identify trends and correlations using statistical software and analysis platforms.[7,8] In contrast, qualitative study seeks to understand underlying consumer motivations and attitudes, often derived from interviews and open-ended survey responses. This two-pronged approach has been instrumental in synthesizing information to derive meaningful insights to inform business strategies.

With technological advances and the rise of machine learning, predictive modeling has become central to marketing strategies. Techniques such as linear regression, logistic regression, and decision trees have enabled marketers to develop models that predict future outcomes and improve understanding of consumer behavior.[9,10] These models have demonstrated considerable effectiveness in digital marketing, allowing companies to tailor their strategies to meet their customers' specific needs and preferences, ultimately leading to increased customer engagement and business growth.[11,12]

Predictive analytics in digital marketing leverages statistical models to forecast consumer behavior, enabling companies to make informed decisions. Using large amounts of digital data, predictive analytics identifies trends and patterns that inform marketing strategies. This approach not only improves understanding of consumer preferences but also helps to optimise resource allocation and improve customer engagement. So, this research aimed to evaluate the use of predictive analytics in digital marketing using a statistical modeling approach to predict consumer behavior.

METHOD

Predictive analytics was applied, using statistical techniques to analyze historical data and forecast future behavior. This process can significantly improve strategic decision-making in digital marketing by providing insights into customer preferences and market trends. The principal methodologies used in predictive analytics included regression analysis, classification models, decision trees, and neural networks, each serving different purposes in the analysis process.

Statistical models evaluated Regression analysis

Regression analysis

Regression analysis is one of the statistical methods used in predictive modeling. It identifies relationships between variables and helps predict outcomes based on historical data. Techniques such as linear regression and logistic regression are particularly valuable for estimating market demand and predicting customer lifetime value. They can also determine how factors such as price and promotional strategies influence sales performance.

Classification models

By evaluating algorithms such as decision trees and support vector machines, marketers can effectively identify high-value customer segments, tailor marketing activities, and optimize advertising targeting. This targeted approach enables personalized marketing strategies that resonate with specific audience segments, ultimately improving customer engagement and conversion rates.

Decision trees

Decision trees were commiserated as intuitive models representing decisions and their possible consequences in a tree-like structure. Each branch signifies a choice, while the leaves represent specific outcomes. This method is beneficial for quick decision-making and can easily handle multiple variables, making it ideal for marketers who must evaluate several factors that influence consumer behavior.

Neural networks

Neural networks were evaluated as they are advanced machine learning models that allow the recognition of complex patterns within large data sets. They are particularly effective at identifying non-linear relationships and can validate predictions made by regression models and decision trees. By employing neural networks, marketers can achieve greater accuracy in predicting consumer behavior and preferences, improving overall marketing effectiveness.

Time series analysis

Time series models were considered within predictive analytics as they analyze data collected over time to identify trends and seasonality. This allows marketers to make informed decisions based on predicted future performance, enabling companies to forecast sales fluctuations and prepare marketing strategies that match expected peak demand.

Data used

The analysis is based on a synthetic dataset that simulates digital consumer behavior, specifically focused on Generation Z (18-24-year-olds). The dataset comprised 1 000 records with the following characteristics:

Demographic Variables

Age: Normally distributed with a mean of 21 years and standard deviation of 2 years, Gender: binary distribution (M/F), Income: log-normal distribution with a base of 10 000 monetary units, Education: categorized into three levels (Secondary 40 %, Tertiary 20 %, Other 40 %) and Region: segmented into five geographical areas (North America, Europe, Asia-Pacific, Latin America, Africa).

Behavioral Variables

Purchase frequency: Poisson distribution with $\lambda=1$ (monthly average), site load time: exponential distribution with 2-unit scale, time on site: exponential distribution with 5-minute scale, ad interactions: uniform distribution between 5 and 10 interactions, conversion rate: binary distribution with 3 % conversion probability, and purchase amount: for converted users, normal distribution with a mean of 75 monetary units and standard deviation of 25.

This dataset simulated a realistic e-commerce scenario, incorporating both demographic and behavioral aspects of the digital consumer, with special emphasis on the youth segment of the market.

RESULTS

In the study 'Predictive Analytics in Digital Marketing: A Statistical Modelling Approach for Predicting Consumer Behaviour,' five machine learning models widely recognized for their effectiveness in classification tasks were compared in detail: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Neural Network. The central purpose was to evaluate how well these models can anticipate consumer behavior on digital platforms, a crucial aspect of strategic decision-making in data-driven marketing campaigns (figure 1).

https://doi.org/10.56294/dm20251061

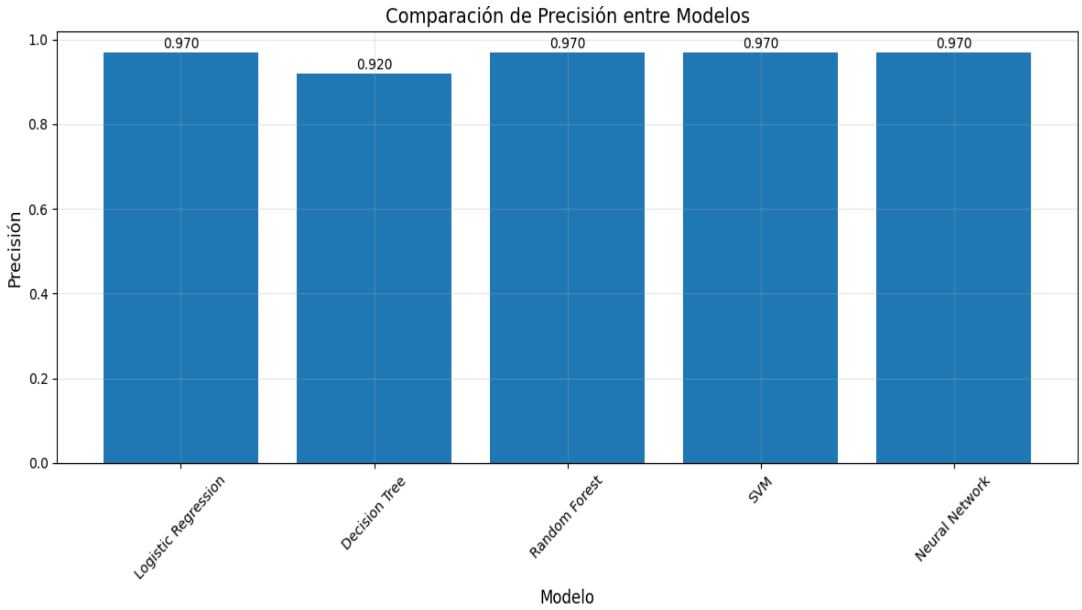

Figure 1. Comparison of accuracy between models

For this purpose, three key indicators were used to assess the model's accuracy and stability over multiple iterations: the overall accuracy (Accuracy), the cross-validation mean (CV Mean), and the cross-validation standard deviation (CV Std). These parameters provide a more complete view of the model's overall performance and its ability to generalize correctly to new data, as seen in figure 2 below.

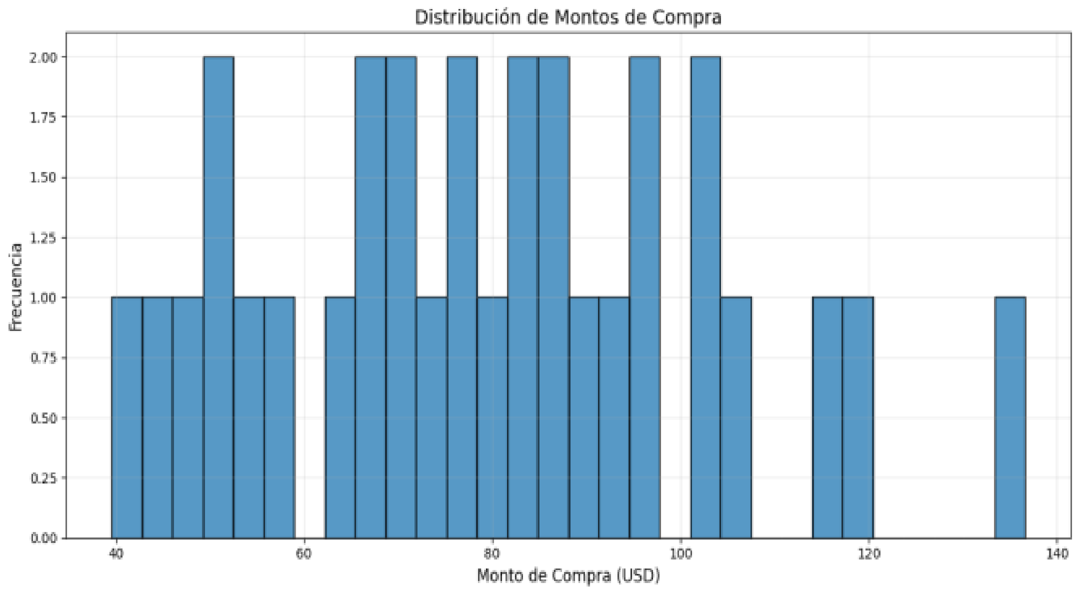

Figure 2. Distribution of purchase amounts

The results obtained were remarkably consistent among four of the five models evaluated. Specifically, logistic regression, random forest, SVM, and neural networks achieved 97 % accuracy in both the test data and the cross-validation mean. Furthermore, they all showed a standard deviation of 0,0, indicating their performance was stable in each cross-validation fold. This uniformity suggests that such models are accurate and reliable to be implemented in real-world contexts where data can be complex and varied, as is the case in digital marketing (figure 3).

In contrast, the Decision Tree performed less well, achieving an accuracy of 92 % and a cross-validation mean slightly above 0,926. Compared to the other models, it showed a standard deviation of approximately 0,0103, evidence of greater variability in its behavior when faced with different data partitions. This sensitivity could be due to its inherent structure, which tends to fit more rigidly to the training data, making it potentially more prone to over-fitting and less effective for modeling more complex or noisy consumption dynamics.

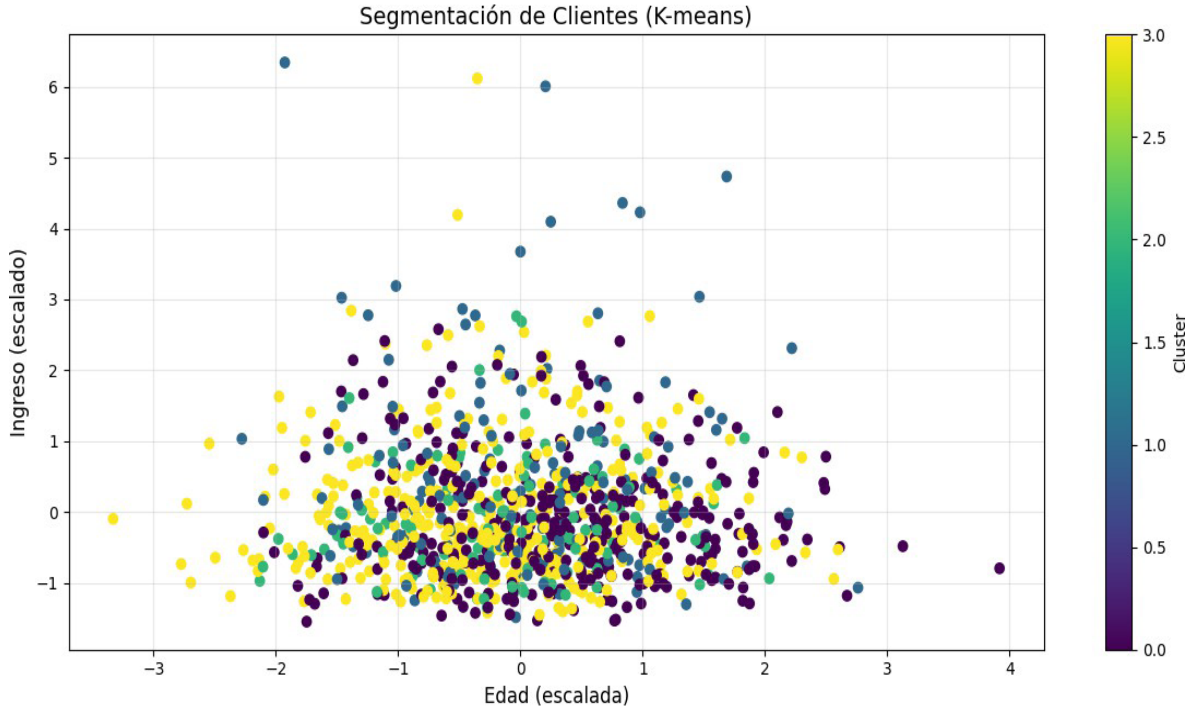

Figure 3. Customer segmentation

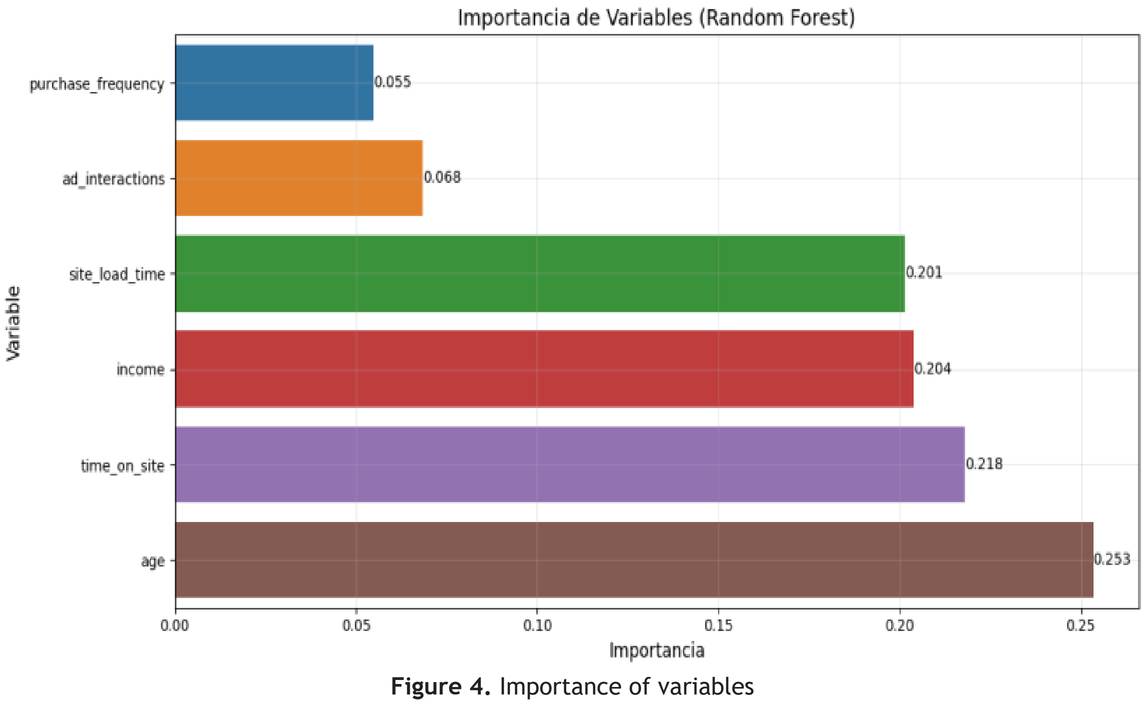

Figure 4. Importance of variables

From an applied perspective, the results highlight the robustness of models such as Random Forest and Neural Network, which maintain high accuracy and consistently do so. This is especially valuable in the digital environment, where small fluctuations in consumer behavior can represent big opportunities or losses for a company. Reliable predictive models allow for more accurate audience targeting, real-time tailoring of advertising messages, and optimization of resources, thereby increasing the effectiveness of advertising campaigns and the profitability of marketing strategies.

Ultimately, this analysis shows that assessing a model's accuracy alone is not enough. It is also crucial to consider its stability, generalisability, and consistent performance across different data samples. These qualities determine the practical value of a predictive model when applied in real and dynamic scenarios such as contemporary digital marketing.

https://doi.org/10.56294/dm20251061

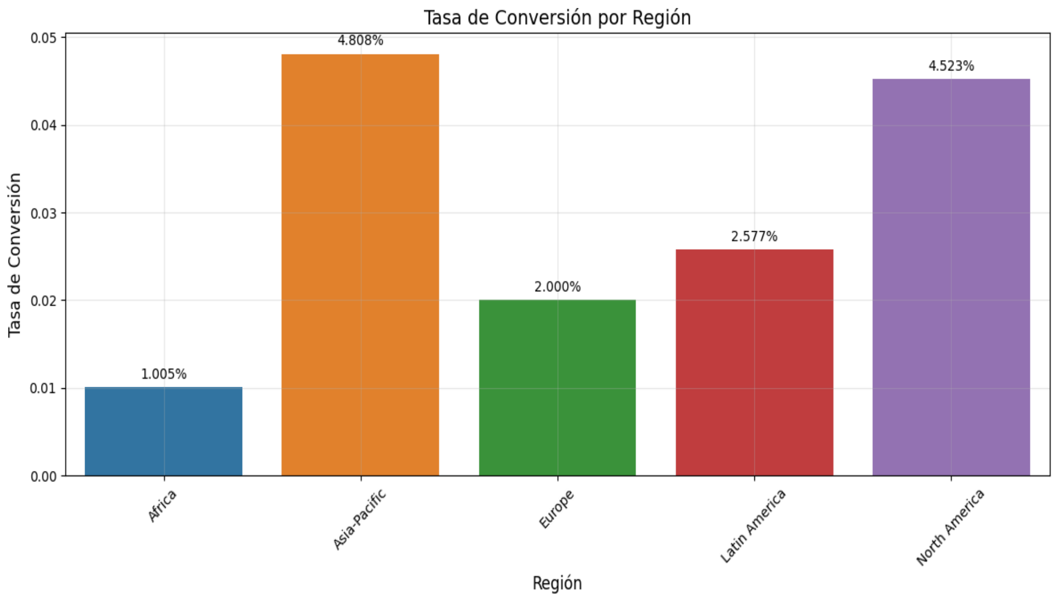

Figure 5. Conversion rate by region

DISCUSSION

The results obtained in this study confirm the potential of predictive analytics as a key tool for anticipating consumer behavior in digital environments. In particular, Logistic Regression, Random Forest, SVM, and Neural Networks models demonstrated outstanding accuracy (97 %) and stability (standard deviation of 0,0), suggesting that they can generalize effectively to new datasets. This consistent predictive ability is critical in the digital marketing domain, where decisions must adapt quickly to market changes and variations in user preferences.[13,14]

These findings are consistent with reports that machine learning-based models have transformed how companies understand and anticipate consumption patterns, significantly improving strategic decision-making.[15] They also point out that model stability is as essential as model accuracy, as it prevents business decisions from being based on random fluctuations or noise in the data.[16]

The underperformance of the Decision Tree, which achieved an accuracy of 92 % and a standard deviation greater than 0,01, highlights a critical aspect of statistical modeling: not all algorithms respond equally well to complex data structures. Its more volatile behavior can be attributed to a lower generalisability and a greater tendency to over-fit.[17,18] While it is an interpretable and easily implemented model, its sensitivity to minimal changes in data may limit its usefulness in dynamic scenarios such as digital marketing, where consistency is key.

From an applied perspective, these results provide academic value and practical implications relevant to marketing teams. For example, implementing robust models such as Neural Networks or Random Forests allows for automating audience segmentation, anticipating the probability of purchase or abandonment, and personalizing messages in real-time. These capabilities align with those that highlight how predictive analytics can improve the customer experience and increase the efficiency of advertising campaigns.[19,20]

Furthermore, considering the role of consumer psychology in designing digital strategies reinforces the idea that predictive models should capture not only transactional data but also behavioral variables, emotional context, and even social signals gathered from digital networks. This is especially important in an era where consumers expect increasingly personalized and relevant interactions.[12,21]

The success of a predictive strategy lies not only in selecting the most accurate model but also in ensuring that it maintains its performance over time and across different consumer segments. This is reflected in the present analysis, where the combination of high accuracy and low variability reinforces the reliability of certain algorithms in real deployment scenarios.[23,24]

Finally, this study highlights the need to integrate predictive analytics tools within automation and campaign management platforms. Platforms that allow constant feedback on model performance and automatic adjustments according to new trends are essential to ensure that predictions remain valid and continue to generate competitive value.[25,26]

In summary, the discussion of these results leads to the conclusion that machine learning models need to be evaluated beyond accuracy metrics in the context of modern digital marketing. Their practical usefulness also depends on their stability, flexibility, and ability to adapt to a constantly evolving digital ecosystem. This multidimensional approach, supported by current literature,[27,28,29] contributes to building more efficient, sustainable, and user-centric predictive systems.

https://doi.org/10.56294/dm20251061

CONCLUSIONS

This study provided evidence of the strategic value of machine learning models applied to predictive analytics in digital marketing, emphasizing predicting consumer behavior. Through the comparison of five widely used models—logistic Regression, Decision Tree, Random Forest, SVM, and Neural Network—it was demonstrated that, beyond accuracy as an isolated indicator, the stability of a model's performance and generalization capacity are determining factors for its effective implementation in real contexts.

Logistic Regression, Random Forest, SVM, and Neural Networks models showed outstanding performance, not only for reaching an accuracy of 97 % but also for maintaining a zero standard deviation, which is evidence of remarkable robustness to variations in the data. These qualities make these algorithms suitable for complex and changing digital environments where decisions must be made quickly and confidently. In contrast, while useful for its interpretability, the Decision Tree showed greater sensitivity to data variability, which could limit its applicability in marketing strategies requiring constant accuracy and adaptive responses.

From a practical perspective, the results of this research reinforce the importance of incorporating predictive solutions into segmentation, campaign personalization, churn analysis, and loyalty processes. These tools allow companies to anticipate their consumers' needs and decisions, optimize resource allocation, and strengthen relationships with their audiences. The true power of predictive analytics in marketing lies not only in the technical capacity of the models but also in how they are integrated into strategic decision-making to generate sustained, customer-centric value. To achieve this, fostering a data-driven organizational culture supported by technological tools and a deep understanding of the human dynamics that define consumption in the digital age is essential.

REFERENCES

1. Theodorakopoulos L, Theodoropoulou A. Leveraging big data analytics for understanding consumer behavior in digital marketing: A systematic review. Hum Behav Emerg Technol. 2024;2024(1):3641502. doi: http://10.1155/2024/3641502.

2. Chkoniya V. Challenges in decoding consumer behavior with data science. Eur J Econ Bus Stud. 2021;7(1):140-53. doi: http://10.26417/897ovg79t.

3. Chaudhary K, Alam M, Al-Rakhami MS, Gumaei A. Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics. J Big Data. 2021;8(1):73. doi: http://10.1186/s40537-021-00466-2.

4. Gan M, Ouyang Y. Study on tourism consumer behavior characteristics based on big data analysis. Front Psychol. 2022;13:876993. doi: http://10.3389/fpsyg.2022.876993.

5. Włodarczyk K. Trends of evolution in consumer behavior in the contemporary world. Probl Zarządzania. 2021;19(1):39-50. doi: http://10.7172/1644-9584.91.3.

6. Gabhane D, Varalaxmi P, Rathod U, Hamida AGB, Anand B. Digital marketing trends: Analyzing the evolution of consumer behavior in the online space. Bol Lit Oral. 2023;10(1):462-73. Available from: https://www.rgcms.edu.in/wp-content/uploads/2023/10/52_TITLE126.pdf

7. Liu H. Big data precision marketing and consumer behavior analysis based on fuzzy clustering and PCA model. J Intell Fuzzy Syst. 2021;40(4):6529-39. doi: http://10.3233/JIFS-189491.

8. Rozenkowska K. Theory of planned behavior in consumer behavior research: A systematic literature review. Int J Consum Stud. 2023;47(6):2670-700. doi: http://10.1111/ijcs.12970.

9. Sunarya PA, Rahardja U, Chen SC, Li YM, Hardini M. Deciphering digital social dynamics: A comparative study of logistic regression and random forest in predicting e-commerce customer behavior. J Appl Data Sci. 2024;5(1):100-13. doi: http://10.47738/jads.v5i1.155.

10. Xiao S, Tong W. Prediction of user consumption behavior data based on the combined model of TF-IDF and logistic regression. J Phys Conf Ser. 2021;1757(1):012089. doi: http://10.1088/1742-6596/1757/1/012089.

11. Mehrolia S, Alagarsamy S, Solaikutty VM. Customers response to online food delivery services during COVID-19 outbreak using binary logistic regression. Int J Consum Stud. 2021;45(3):396-408. doi: http://10.1111/ijcs.12630.

12. Arumugam T, Hameed SS, Sanjeev MA. Buyer behaviour modelling of rural online purchase intention using logistic regression. Int J Manag Enterp Dev. 2023;22(2):139-57. doi: http://10.1504/IJMED.2023.130153.

13. Kasem MS, Hamada M, Taj-Eddin I. Customer profiling, segmentation, and sales prediction using AI in direct marketing. Neural Comput Appl. 2024;36(9):4995-5005. doi: http://10.1007/s00521-023-09339-6.

14. Habil S, El-Deeb S, El-Bassiouny N. AI-based recommendation systems: the ultimate solution for market prediction and targeting. In: The Palgrave handbook of interactive marketing. Cham: Springer; 2023. p. 683-704. Available from: https://www.researchgate.net/publication/388871670_AI_and_Machine_Learning_in_Multimedia_Platforms_Present_Applications_and_Future_Implications

15. Chakravarthy BS, Perlmutter HV. Strategic planning for a global business. Int Strateg Manag. 2023:29-42. doi: http://10.4324/9781315058993-4.

16. Zhang H, Zang Z, Zhu H, Uddin MI, Amin MA. Big data-assisted social media analytics for business model for business decision making system competitive analysis. Inf Process Manag. 2022;59(1):102762. doi: http://10.1016/j.ipm.2021.102762.

17. Hu J, Szymczak S. A review on longitudinal data analysis with random forest. Brief Bioinform. 2023;24(2):bbad002. doi: http://10.1093/bib/bbad002.

18. Sun Z, Wang G, Li P, Wang H, Zhang M, Liang X. An improved random forest based on the classification accuracy and correlation measurement of decision trees. Expert Syst Appl. 2024;237:121549. doi: http://10.1016/j.eswa.2023.121549.

19. De Araujo Morais LR, da Silva Gomes GS. Neural Network-Enhanced Decision Support: Investigating Prediction Intervals for Real-Time Digital Marketing Return on Investment Data. In: Brazilian Workshop on Social Network Analysis and Mining (BraSNAM). SBC; 2024. p. 47-60. doi: http://10.5753/brasnam.2024.2232.

20. Mejía Trejo J. Digital marketing model innovation and Generation Z as consumer decision-making style. Making predictions with an artificial neural network, in Mexico. Contad Adm. 2021;66(4). doi: http://10.22201/fca.24488410e.2021.3258.

21. Ferreira JJ, Fernandes CI, Rammal HG, Veiga PM. Wearable technology and consumer interaction: A systematic review and research agenda. Comput Hum Behav. 2021;118:106710. doi: http://10.1016/j.chb.2021.106710.

22. Mogaji E, Balakrishnan J, Nwoba AC, Nguyen NP. Emerging-market consumers' interactions with banking chatbots. Telemat Inform. 2021;65:101711. doi: http://10.1016/j.tele.2021.101711.

23. Khan M, Ahmad M, Alidjonovich RD, Bakhritdinovich KM, Turobjonovna KM, Odilovich IJ. The impact of cultural factors on digital marketing strategies with Machine learning and honey bee Algorithm (HBA). Cogent Bus Manag. 2025;12(1):2486590. doi: http://10.1080/23311975.2025.2486590.

24. Wang G. Customer segmentation in the digital marketing using a Q-learning based differential evolution algorithm integrated with K-means clustering. PLoS One. 2025;20(2):e0318519. doi: http://10.1371/journal.pone.0318519.

25. Kasem MS, Hamada M, Taj-Eddin I. Customer profiling, segmentation, and sales prediction using AI in direct marketing. Neural Comput Appl. 2024;36(9):4995-5005. doi: http://10.1007/s00521-023-09339-6.

26. Sharma A, Fadahunsi A, Abbas H, Pathak VK. A multi-analytic approach to predict social media marketing influence on consumer purchase intention. J Indian Bus Res. 2022;14(2):125-49. doi: http://10.1108/JIBR-08-2021-0313.

27. Ullal MS, Hawaldar IT, Soni R, Nadeem M. The role of machine learning in digital marketing. SAGE Open. 2021;11(4):21582440211050394. doi: http://10.1177/21582440211050394.

28. Berlilana B, Hariguna T, El Emary IM. Enhancing Digital Marketing Strategies with Machine Learning for

Analyzing Key Drivers of Online Advertising Performance. J Appl Data Sci. 2025;6(2):817-27. doi: http://10.47738/jads.v6i2.658.

29. Gutnik S. Application of data mining and machine learning methods to enhance the effectiveness of digital marketing strategies. In: Digital Strategies in a Global Market: Navigating the Fourth Industrial Revolution. 2021. p. 131-44. doi: http://10.1007/978-3-030-58267-8_10.

FUNDING
The authors received no funding for the conduct of this research.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION
Conceptualisation: Jazmín Isabel García-Guerra.
Data Curation: Jazmín Isabel García-Guerra.
Formal analysis: Héctor Aguilar-Cajas.
Research: Héctor Aguilar-Cajas.
Methodology: Héctor Aguilar-Cajas.
Project administration: Héctor Aguilar-Cajas.
Resources: Ana Lucía Rivera-Abarca.
Software: Ana Lucía Rivera-Abarca.
Supervision: Ana Lucía Rivera-Abarca.
Validation: Jazmín Isabel García-Guerra.
Visualisation: Jazmín Isabel García-Guerra.
Writing - original draft: Freddy Armijos-Arcos and José Israel López-Pumalema.
Writing - revision and editing: Freddy Armijos-Arcos and José Israel López-Pumalema.