International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 08, Issue – 06, June 2023, PP – 47-51

Evaluating Machine Learning Models for Engagement Prediction in Different Photograph Types

Alan Galván Espinoza[1], Silvestre Guillermo Puebla Serrano[2], Armida Gonzales Lorence[3]

[1]*Administrative Engineering,Tecnológico Nacional de México/Instituto Tecnológico de Cuautla, México*
[2]*Administrative Engineering, Tecnológico Nacional de México/Instituto Tecnológico de Cuautla, México*
[3]*Administrative Engineering, Tecnológico Nacional de México/Instituto Tecnológico de San Juan Del Río, México*

Abstract: This study evaluates different machine learning models for predicting engagement in photographs. The models were loaded into the Orange platform, known for its visual machine learning capabilities. Labeled datasets were collected, including information on the photograph type and engagement level. Various models, including decision trees, logistic regression, and neural networks, were constructed and trained. Performance evaluation involved cross-validation techniques and comparison of metrics such as accuracy and cross-correlation. The model with the best performance was selected to predict engagement in new posts based on photograph type.

The results highlight the importance of selecting the appropriate model for specific project objectives. The comparative analysis revealed no standout model, but valuable insights were obtained regarding the types of photographs that can be predicted with higher precision. The study's systematic evaluation approach helps identify model strengths and weaknesses and uncover hidden patterns in the database. The findings demonstrate the need for improving the prediction capacity of machine learning models in engagement prediction. This research contributes to a better understanding of engagement prediction in photographs and emphasizes the significance of model selection and dataset characteristics in achieving accurate predictions.

Keywords: Engagement prediction, Machine learning models, Photograph type, Performance evaluation, Model selection

I. INTRODUCTION

The In today's digital era, effective marketing strategies and data analysis techniques play a vital role in driving business success. With the widespread adoption of online platforms and social media, companies have access to vast amounts of data that can be harnessed to understand customer behavior, preferences, and engagement patterns. Leveraging this data effectively can enable businesses to tailor their marketing efforts, optimize customer experiences, and ultimately enhance their overall competitiveness.

One of the key challenges in digital marketing is predicting and improving customer engagement. Engaged customers are more likely to interact with brands, make purchases, and become loyal advocates. However, accurately predicting and understanding engagement levels can be complex due to the dynamic and multifaceted nature of customer interactions. This is where machine learning techniques, such as those employed in data analysis, play a crucial role.

Previous research has explored various approaches to analyze customer engagement in the digital realm. Traditional statistical models have been employed, but they often struggle to capture the intricacies of dynamic online environments. With the advent of machine learning, the ability to process and analyze large datasets has significantly improved, allowing for more accurate predictions and insights into customer behavior.

II. METHODS

To assess the efficiency of different engagement prediction models based on the type of photograph, the models were loaded into Orange due to its ease of use as a visual machine learning platform. First, the input data was prepared by collecting and selecting a labeled dataset that included information about the type of photograph and the level of engagement obtained in each post.

Next, various machine learning models were constructed and trained, including decision trees, logistic regression, and neural networks. For model evaluation (Figure 30), different cross-validation techniques were employed, and performance metrics such as accuracy and cross-correlation were compared. Finally, the model with the best performance was selected to make predictions about the engagement of new posts based on the type of photograph.

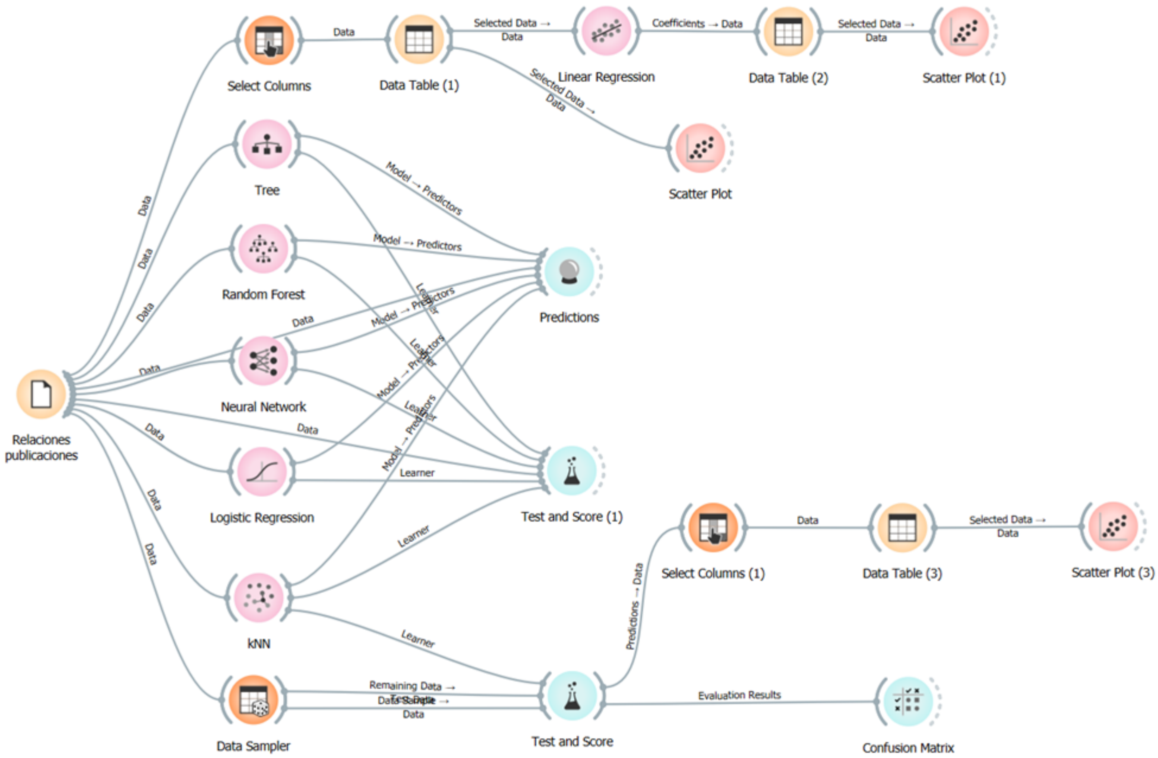
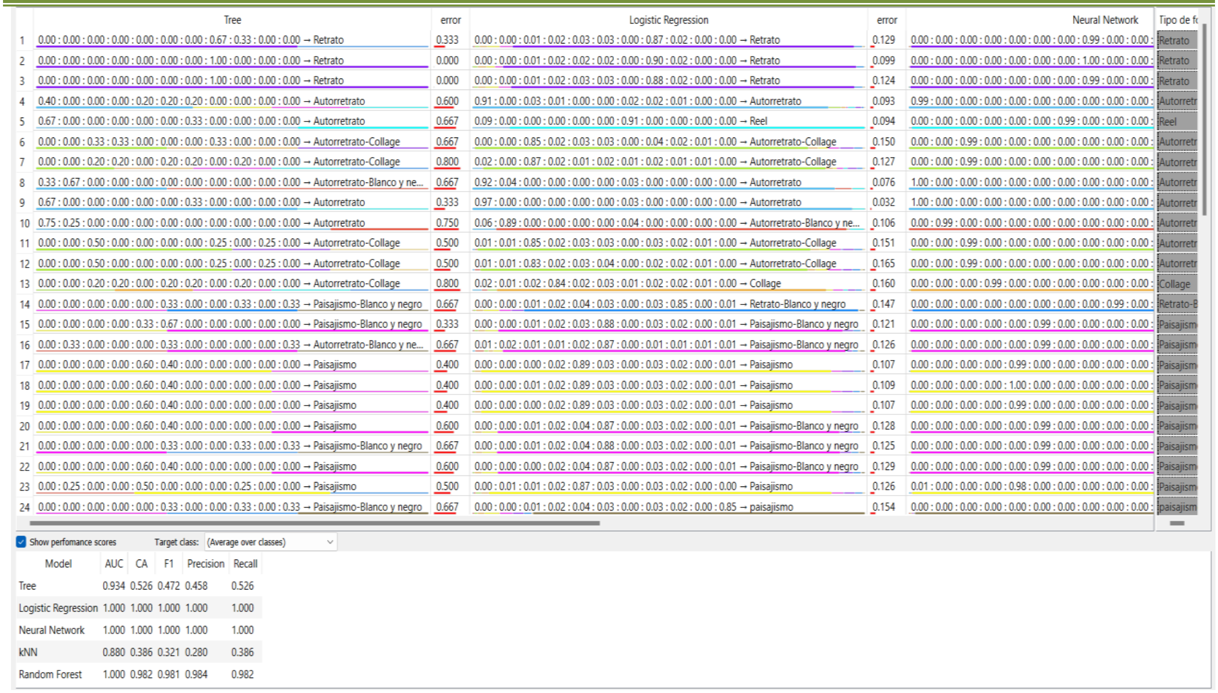


Figure 30 Data flow for the evaluation and prediction of multiple machine learning models.

During the evaluation of different machine learning models in engagement prediction, a file containing a database was used, which passed through various nodes pre-loaded with common machine learning models. These models were compared in prediction and testing nodes to determine which one offered the best accuracy in engagement prediction.

Multiple types of photographs were evaluated using each machine learning model in each run, allowing for a broader and more comprehensive understanding of the performance of each model in different contexts and situations (Figure 31). Through these results, it was possible to determine which models provided the highest accuracy and which types of photographs were more likely to be predicted with greater precision.

This systematic and rigorous evaluation approach of multiple machine learning models in engagement prediction allows for identifying the strengths and weaknesses of each model and using this information to improve prediction accuracy in the future. Additionally, it helps identify patterns and trends in the database that might otherwise go unnoticed.



	Tree	error	Logistic Regression	error	Neural Network	Tipo de fo
1	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.67 : 0.33 : 0.00 : 0.00 → Retrato	0.333	0.00 : 0.00 : 0.01 : 0.02 : 0.03 : 0.03 : 0.00 : 0.87 : 0.02 : 0.00 : 0.00 → Retrato	0.129	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.99 : 0.00 : 0.00 :	Retrato
2	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 1.00 : 0.00 : 0.00 : 0.00 → Retrato	0.000	0.00 : 0.00 : 0.01 : 0.02 : 0.02 : 0.02 : 0.00 : 0.90 : 0.02 : 0.00 : 0.00 → Retrato	0.099	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 1.00 : 0.00 : 0.00 :	Retrato
3	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 1.00 : 0.00 : 0.00 : 0.00 → Retrato	0.000	0.00 : 0.00 : 0.01 : 0.02 : 0.03 : 0.03 : 0.00 : 0.88 : 0.02 : 0.00 : 0.00 → Retrato	0.124	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.99 : 0.00 : 0.00 :	Retrato
4	0.40 : 0.00 : 0.00 : 0.00 : 0.20 : 0.20 : 0.20 : 0.00 : 0.00 : 0.00 : 0.00 → Autorretrato	0.600	0.91 : 0.00 : 0.03 : 0.01 : 0.00 : 0.00 : 0.02 : 0.02 : 0.01 : 0.00 : 0.00 → Autorretrato	0.093	0.99 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 :	Autorretr
5	0.67 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.33 : 0.00 : 0.00 : 0.00 : 0.00 → Autorretrato	0.667	0.09 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.91 : 0.00 : 0.00 : 0.00 : 0.00 → Reel	0.094	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.99 : 0.00 : 0.00 : 0.00 :	Reel
6	0.00 : 0.00 : 0.33 : 0.33 : 0.00 : 0.00 : 0.00 : 0.33 : 0.00 : 0.00 : 0.00 → Autorretrato-Collage	0.667	0.00 : 0.00 : 0.85 : 0.02 : 0.03 : 0.03 : 0.00 : 0.04 : 0.02 : 0.01 : 0.00 → Autorretrato-Collage	0.150	0.00 : 0.00 : 0.99 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 :	Autorretr
7	0.00 : 0.00 : 0.20 : 0.20 : 0.00 : 0.20 : 0.20 : 0.00 : 0.20 : 0.00 : 0.00 → Autorretrato-Collage	0.800	0.02 : 0.00 : 0.87 : 0.02 : 0.01 : 0.02 : 0.01 : 0.02 : 0.01 : 0.01 : 0.00 → Autorretrato-Collage	0.127	0.00 : 0.00 : 0.99 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 :	Autorretr
8	0.33 : 0.67 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 → Autorretrato-Blanco y ne...	0.667	0.92 : 0.04 : 0.00 : 0.00 : 0.00 : 0.00 : 0.03 : 0.00 : 0.00 : 0.00 : 0.00 → Autorretrato	0.076	1.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 :	Autorretr
9	0.67 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.33 : 0.00 : 0.00 : 0.00 : 0.00 → Autorretrato	0.333	0.97 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.03 : 0.00 : 0.00 : 0.00 : 0.00 → Autorretrato	0.032	1.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 :	Autorretr
10	0.75 : 0.25 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 → Autorretrato	0.750	0.06 : 0.89 : 0.00 : 0.00 : 0.00 : 0.00 : 0.04 : 0.00 : 0.00 : 0.00 : 0.00 → Autorretrato-Blanco y ne...	0.106	0.00 : 0.99 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 :	Autorretr
11	0.00 : 0.00 : 0.50 : 0.00 : 0.00 : 0.00 : 0.00 : 0.25 : 0.00 : 0.25 : 0.00 → Autorretrato-Collage	0.500	0.01 : 0.01 : 0.85 : 0.02 : 0.03 : 0.03 : 0.00 : 0.03 : 0.02 : 0.01 : 0.00 → Autorretrato-Collage	0.151	0.00 : 0.00 : 0.99 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 :	Autorretr
12	0.00 : 0.00 : 0.50 : 0.00 : 0.00 : 0.00 : 0.00 : 0.25 : 0.00 : 0.25 : 0.00 → Autorretrato-Collage	0.500	0.01 : 0.01 : 0.83 : 0.02 : 0.03 : 0.04 : 0.00 : 0.02 : 0.02 : 0.01 : 0.00 → Autorretrato-Collage	0.165	0.00 : 0.00 : 0.99 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 :	Autorretr
13	0.00 : 0.00 : 0.20 : 0.20 : 0.00 : 0.20 : 0.20 : 0.00 : 0.20 : 0.00 : 0.00 → Autorretrato-Collage	0.800	0.02 : 0.01 : 0.02 : 0.84 : 0.02 : 0.03 : 0.01 : 0.02 : 0.02 : 0.01 : 0.00 → Collage	0.160	0.00 : 0.00 : 0.00 : 0.99 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 :	Collage
14	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.33 : 0.00 : 0.00 : 0.33 : 0.00 : 0.33 → Paisajismo-Blanco y negro	0.667	0.00 : 0.00 : 0.01 : 0.02 : 0.04 : 0.03 : 0.00 : 0.03 : 0.85 : 0.00 : 0.01 → Retrato-Blanco y negro	0.147	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.99 : 0.00 :	Retrato-B
15	0.00 : 0.00 : 0.00 : 0.00 : 0.33 : 0.67 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 → Paisajismo-Blanco y negro	0.333	0.00 : 0.00 : 0.01 : 0.02 : 0.03 : 0.88 : 0.00 : 0.03 : 0.02 : 0.00 : 0.01 → Paisajismo-Blanco y negro	0.121	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.99 : 0.00 : 0.00 : 0.00 : 0.00 :	Paisajism
16	0.00 : 0.33 : 0.00 : 0.00 : 0.00 : 0.33 : 0.00 : 0.00 : 0.00 : 0.00 : 0.33 → Autorretrato-Blanco y ne...	0.667	0.01 : 0.02 : 0.01 : 0.01 : 0.02 : 0.87 : 0.00 : 0.01 : 0.01 : 0.01 : 0.01 → Paisajismo-Blanco y negro	0.126	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.99 : 0.00 : 0.00 : 0.00 : 0.00 :	Paisajism
17	0.00 : 0.00 : 0.00 : 0.00 : 0.60 : 0.40 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 → Paisajismo	0.400	0.00 : 0.00 : 0.00 : 0.02 : 0.89 : 0.03 : 0.00 : 0.03 : 0.02 : 0.00 : 0.01 → Paisajismo	0.107	0.00 : 0.00 : 0.00 : 0.00 : 0.99 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 :	Paisajism
18	0.00 : 0.00 : 0.00 : 0.00 : 0.60 : 0.40 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 → Paisajismo	0.400	0.00 : 0.00 : 0.01 : 0.02 : 0.89 : 0.03 : 0.00 : 0.03 : 0.02 : 0.00 : 0.01 → Paisajismo	0.109	0.00 : 0.00 : 0.00 : 0.00 : 1.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 :	Paisajism
19	0.00 : 0.00 : 0.00 : 0.00 : 0.60 : 0.40 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 → Paisajismo	0.400	0.00 : 0.00 : 0.00 : 0.02 : 0.89 : 0.03 : 0.00 : 0.03 : 0.02 : 0.00 : 0.01 → Paisajismo	0.107	0.00 : 0.00 : 0.00 : 0.00 : 0.99 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 :	Paisajism
20	0.00 : 0.00 : 0.00 : 0.00 : 0.60 : 0.40 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 → Paisajismo	0.600	0.00 : 0.00 : 0.01 : 0.02 : 0.04 : 0.87 : 0.00 : 0.03 : 0.02 : 0.00 : 0.01 → Paisajismo-Blanco y negro	0.128	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.99 : 0.00 : 0.00 : 0.00 : 0.00 :	Paisajism
21	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.33 : 0.00 : 0.00 : 0.33 : 0.00 : 0.33 → Paisajismo-Blanco y negro	0.667	0.00 : 0.00 : 0.01 : 0.02 : 0.04 : 0.88 : 0.00 : 0.03 : 0.02 : 0.00 : 0.01 → Paisajismo-Blanco y negro	0.125	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.99 : 0.00 : 0.00 : 0.00 : 0.00 :	Paisajism
22	0.00 : 0.00 : 0.00 : 0.00 : 0.60 : 0.40 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 → Paisajismo	0.600	0.00 : 0.00 : 0.00 : 0.02 : 0.04 : 0.87 : 0.00 : 0.03 : 0.02 : 0.00 : 0.01 → Paisajismo-Blanco y negro	0.129	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.99 : 0.00 : 0.00 : 0.00 : 0.00 :	Paisajism
23	0.00 : 0.25 : 0.00 : 0.00 : 0.50 : 0.00 : 0.00 : 0.00 : 0.25 : 0.00 : 0.00 → Paisajismo	0.500	0.00 : 0.01 : 0.01 : 0.02 : 0.87 : 0.03 : 0.00 : 0.03 : 0.02 : 0.00 : 0.00 → Paisajismo	0.126	0.01 : 0.00 : 0.00 : 0.00 : 0.98 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 :	Paisajism
24	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.33 : 0.00 : 0.00 : 0.33 : 0.00 : 0.33 → Paisajismo-Blanco y negro	0.667	0.00 : 0.00 : 0.01 : 0.02 : 0.04 : 0.03 : 0.00 : 0.03 : 0.02 : 0.00 : 0.85 → paisajismo	0.154	0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 : 0.00 :	paisajism

Show perfomance scores Target class: (Average over classes)

Model	AUC	CA	F1	Precision	Recall
Tree	0.934	0.526	0.472	0.458	0.526
Logistic Regression	1.000	1.000	1.000	1.000	1.000
Neural Network	1.000	1.000	1.000	1.000	1.000
kNN	0.880	0.386	0.321	0.280	0.386
Random Forest	1.000	0.982	0.981	0.984	0.982

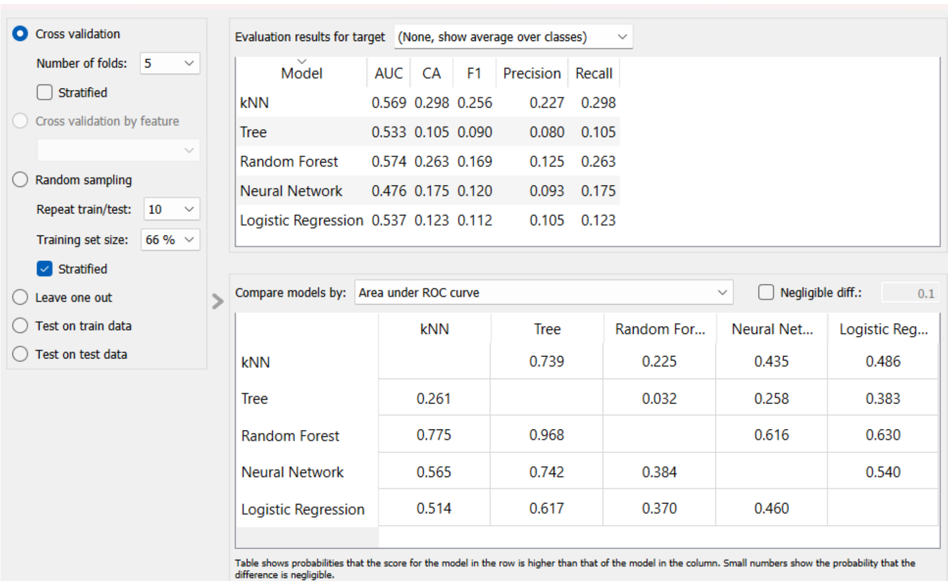
Figure 31 Tests for each model in each variable and result matrix with their differences.

The comparative matrix among the models enables evaluating their performance in terms of different indicators such as AUC (area under the ROC curve), accuracy, recall, and F1 score. These indicators allow determining which model is most suitable for predicting user engagement with different photographs.

III. RESULTS

The comparative matrix (Figure 32) displays the performance of each model for each type of photograph. By comparing the results, we can identify which model performs best for each type of photograph. Furthermore, we can identify models that consistently perform well across all evaluated types of photographs.

It is important to consider that choosing the appropriate model depends on the specific objectives and requirements of the project. Therefore, carefully evaluating the results from the comparative matrix and selecting the model that best fits the project's needs is crucial.



Model	AUC	CA	F1	Precision	Recall
kNN	0.569	0.298	0.256	0.227	0.298
Tree	0.533	0.105	0.090	0.080	0.105
Random Forest	0.574	0.263	0.169	0.125	0.263
Neural Network	0.476	0.175	0.120	0.093	0.175
Logistic Regression	0.537	0.123	0.112	0.105	0.123

Compare models by: Area under ROC curve Negligible diff.: 0.1

	kNN	Tree	Random For...	Neural Net...	Logistic Reg...
kNN		0.739	0.225	0.435	0.486
Tree	0.261		0.032	0.258	0.383
Random Forest	0.775	0.968		0.616	0.630
Neural Network	0.565	0.742	0.384		0.540
Logistic Regression	0.514	0.617	0.370	0.460	

Table shows probabilities that the score for the model in the row is higher than that of the model in the column. Small numbers show the probability that the difference is negligible.

Figure 32 Detailed review of the differences between models

Based on the results obtained from the evaluated machine learning models, it can be concluded that no model stands out among the others in predicting engagement in photographs. The model with the best performance was Random Forest with an AUC of 0.574, but it still did not achieve a satisfactory level of prediction. Additionally, the values of CA, F1, Precision, and Recall for all models were relatively low, indicating ample room for improving the prediction capacity of these models. The results may be influenced by the amount of information available in the database.

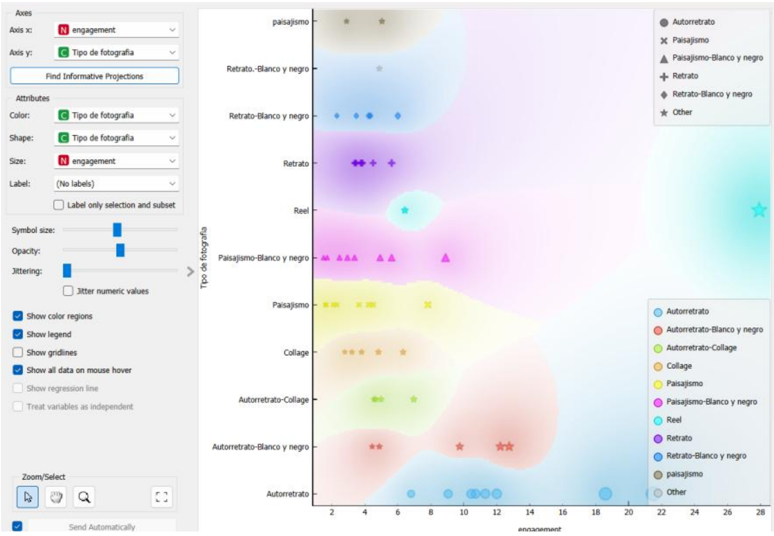


Figure 33 Segmentation graph by type of post

In the above graph, we can observe the distribution of different types of photographs and their level of engagement (Figure 33). Reels stand out as the least generated content type but potentially with higher engagement. Self-portraits consistently show high levels of engagement (Figure 34).

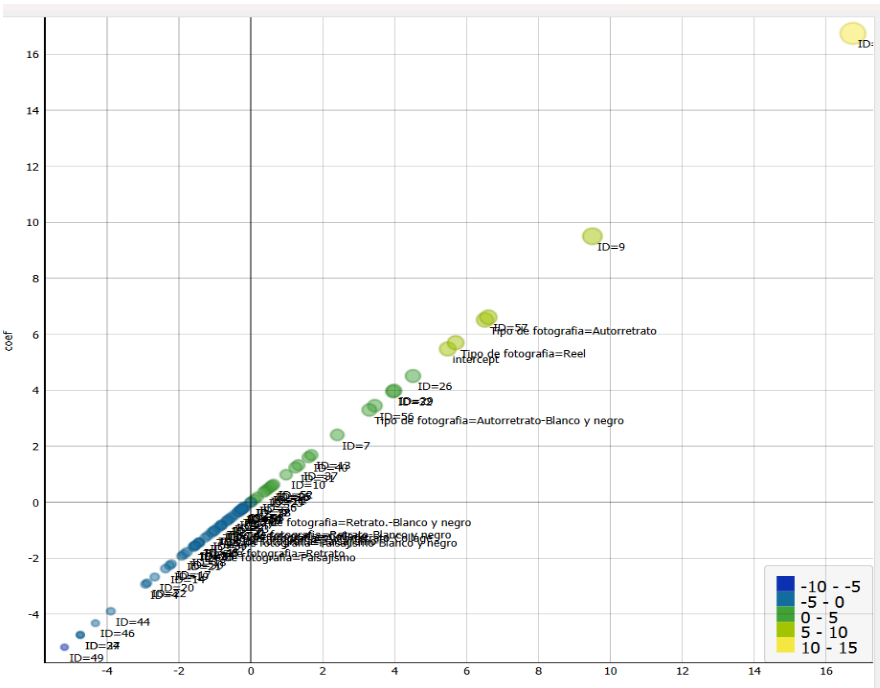


Figure 34 Linear regression graph of the studied photographic categories

In summary, through the systematic evaluation of different machine learning models in predicting engagement in photographs, it has been identified that none of the evaluated models offer outstanding performance. However, valuable insights have been gained regarding the types of photographs that are more likely tobe predicted with higher accuracy. It has also emphasized the importance of selecting the appropriate model to meet specific project objectives. Furthermore, the distribution graph of different types of photographs and their engagement levels has revealed that reels and self-portraits are content types that can generate considerably high engagement responses.

By conducting this evaluation, we have gathered valuable information that can guide future improvements in prediction accuracy. Although none of the models achieved exceptional performance, the insights gained from this study will contribute to refining the prediction models and enhancing their effectiveness.

IV. CONCLUSION

In conclusion, the evaluation of various machine learning models in predicting engagement in photographs has provided valuable findings. While no model stands out as superior, the analysis has shed light on the types of photographs that are more likely to yield accurate predictions. These insights, along with the comparative matrix and distribution graphs, will inform the selection of appropriate models and help optimize the prediction of engagement in future projects.

ACKNOWLEDGEMENTS

During the development of this thesis, I had the opportunity to work in three main areas that form the foundation of this work: statistics, marketing, and machine learning. Although I was able to study their fundamentals, I discovered that multiple variables affecting the results were encountered when working with them, ranging from relevant input data to the implementation of machine learning models with their diverse functions and applications in marketing strategies.

The use of these tools in the real world is an ongoing process of field testing, as the results can be highly fluctuating. What works today may not work tomorrow, so we can only speculate about it.

Throughout this work, I have learned that the combination of these areas is essential for making informed and effective decisions. The interaction between them not only allows us to achieve better results but also to develop new solutions and approaches for upcoming social, economic, and technological challenges.

REFERENCES

Books:

[1]. Orlando Troisi, G. M. (2019). Growth hacking: Insights on data-driven decision-making from three firms. Roma: Elsevier.

Chapters in Books:

[2]. Muhammad Ridwan Andi Purnomo, A. A. (2020). Effective Marketing Strategy Determination Based on Customers Clustering Using Machine Lerning Technique. Islam indonesia: Journal of physics: conference series.

Theses:

[3]. Maisueche Cuadrado, A. (2019). Utilizacióndel Machine Learning en la industria 4.0. Valladolid: Universidad de valladolid, Escuela de ingenierías industriales.

[4]. Saura, J. R. (2020). Using data science in digital marketing: Framework, methods, and performance metrics. Madrid España: Journal of innovation & knowledge.

[5]. Matich, D. J. (2001). Redes Neuronales: Conceptos Básicos y Aplicaciones. Regional Rosario: Universidad Tecnológica Nacional.

