Customer churn prediction in e-commerce base using machine learning and LIME algorithm

Eduarda Neves da Silva
Escola Politécnica
Universidade de São Paulo
São Paulo, Brasil
eduarda.neves@usp.br

Filipe Bento Magalhaes
Departamento da Engenharia
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo
Cubatão, Brazil
filipem@lme.usp.br

Walter Jaimes Salcedo
Escola Politécnica
Universidade de São Paulo
São Paulo, Brazil
walter.salcedo@usp.br

***Abstract*—Anticipating customer churn trends is extremely important in the current competitive scenario. This allows the company to take preventive measures before churn occurs, as acquiring a new customer can cost up to five times more than retaining an existing customer. Furthermore, understanding the reasons that lead to churn in different scenarios is essential to creating an effective strategy to meet each customer's needs. The present work aims to apply a methodology that combines machine learnin g techniques to enable the anticipation of churn trends and group customers with common motivations. The methodology applied consisted of the following steps: developing a predictive churn model, applying the LIME (Local Interpretable Model-Agnostic Explanations) interpretability technique to identify the most influential characteristics in decisions in prone customers and, finally, using K-means clustering to group customers according to their common motivations. This methodology was successfully applied on an e-commerce basis, reaching an area under curve parameter (AUC) of 78% and the creation of 5 final clusters. This approach makes a methodological contribution to future studies and practical experiments in the field of churn analysis.**

Keywords—Predict model, K-means, Churn, Analytical Models, Machine Learning, Clustering

I. Introduction

In the realm of business, 'churn' denotes the frequency at which customers discontinue their utilization of a company's services or products within a specified timeframe. Within today's fiercely competitive landscape, the e-commerce sector grapples with elevated rates of customer churn, highlighting the paramount importance of predictive analytics and a deep understanding of customer behavior for sustaining corporate vitality [1]. Studies underscore that the expenditure required to retain an existing customer pales in comparison—ranging from five to ten times less—against the costs associated with acquiring new clientele [2]. Furthermore, fostering customer satisfaction and cultivating loyalty emerge as pivotal drivers for increasing company revenues and fortifying profitability [3]. Several studies have emerged on the topic, in which the application of machine learning techniques to predict churn is widely used [4], [5], [6], [7]. However, understanding the motivations that lead customers to churn in different scenarios can be complex and require specific data analysis techniques, given the plurality of customers [8]. Failure to understand these individual needs and causes can lead to inadequate communications and suggestions that are not relevant, making it difficult to create an effective strategy that reaches the specificities of the target audience. Therefore, churn propensity may be an important metric, but it is not the only consideration to be made when dealing with consumers. It is necessary to understand the different reasons for each, so that appropriate measures can be taken to prevent their loss [9]. Aiming to advance theoretical and methodological knowledge in the development of churn analysis strategies, this work aims to apply the following machine learning techniques:

- Develop a predictive model to predict customer churn propensity on e-commerce platforms.
- Apply LIME (Local Interpretable Model-Agnostic Explanations) to interpret the predictive model and identify the most influential characteristics on propensity.
- Use the K-means clustering algorithm to segment customers based on the characteristics identified by LIME.

In the present work we propose a methodology to develop models for predicting customer turnover and identify the reasons for this effect, in order to assist in the generation of inputs for the development of more assertive retention strategies, through the clustering techniques associated to predictive models. The proposed method was inspired by the work of Sun, Sharma and Mat (2023) who developed a model for detecting and grouping patients with cardiovascular problems, supporting the generation of inputs for an assertive diagnosis [10].

II. Literature Review

Machine Learning is a subfield of Artificial Intelligence (AI) that uses data as input and, through algorithms and statistical models, predicts new results. Some literature divides machine learning into four types: Supervised, Unsupervised, Semi-supervised and Reinforcement. In this work, supervised and unsupervised learning techniques were combined for prediction, detection and grouping of customer profiles. Supervised learning is where the model is fed labeled input and output data, with the aim of learning to predict the output for new input data. Unsupervised learning is used to discover subjective patterns in data that do not have prior labels [11].

A. Supervised Learning

To develop the predictive model, four supervised learning algorithms were applied to compare performances: logistic regression, random forest (RF), Gradient Boosting Machine (GBM) and Light Gradient Boosting Machine (LightGBM). All techniques used are focused on solving the binary classification problem.

Equation 1 presents the basis of the Logistic Regression solution, which uses a sigmoid function to predict the output

2024 IEEE ANDESCON | 979-8-3503-5528-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ANDESCON61840.2024.10755666

979-8-3503-5528-4/24/$31.00 ©2024 IEEE

"y" as a categorical dependent variable using independent variables x_n [12].

$$P(Y = 1) = \frac{1}{1 + e^{-g(x)}} \qquad (1)$$

Where g(x) is:

$$g(x) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \qquad (2)$$

The algorithm works to estimate the values of α and β in order to minimize the Loss Function L obtained between the real value and the predicted value.

Another method used for prediction was Random Forest, which is based on the principle of ensemble learning that combines several classifiers to solve complex problems and increase the model's accuracy. Its operation is based on the use of Decision Tree (DT). DTs are machine learning models based on a tree-like flow diagram, representing all possibilities and outcomes of a data set. RF works by creating several random decision trees, in which each tree is built with a random subset of the training data and model variables. To make a prediction for a new sample, each tree in the forest makes its own prediction and RF considers the result based on the majority of votes to improve accuracy [12].

Unlike RF, which creates uncorrelated trees, Gradient Boosting consists of correlated trees that are created iteratively from the prediction errors of the previous tree to minimize the final prediction error. In each iteration, the model is fine-tuned to the training data, where incorrectly classified examples are weighted more heavily [13]. Generally, GBM starts its iteration with a constant value and then in the new model corrects the error of the previous one by minimizing the Loss Function $L(y_i, F_m(x_i))$. Equation 3 describes the final GBM model using a data set $\{(x_i, y_i)\}$:

$$F_M(x) = F_0(x) + \sum_{m=1}^{M} F_m(x) \qquad (3)$$

The $F_m(x)$ are functions constructed in a stagewise, the model is constructed by adding estimators incrementally and sequentially, as opposed to methods such as logistic regression, where all parameters are adjusted simultaneously. These functions are parameterized as $\beta_m h(x;\alpha_m)$, where β_m is a weight, and α_m the parameters of the learner h. Therefore, the algorithm works to find all the optimal values of β_m and α_m that minimize the Loss Function $L(y_i, F_m(x_i))$ for each m=1,2,…,M, according to expressions (4), and (5) in their vectorized notation.

$$(\beta_m, \alpha_m) = \arg\min_{\beta,\alpha} \sum_{i=1}^{n} L(y^{(i)}, F_{m-1} x^{(i)} + \beta h(x^{(i)}; \alpha)) \qquad (4)$$

$$F_m(\mathbf{X}) = F_{m-1}(\mathbf{X}) + \eta \Delta_m(X) \qquad (5)$$

The last prediction is called pseudo-residual and defined as r_{m-1}, which is used to adjust the learner $h(x;\alpha_m)$ with weight β:

$$r_{m_1} = \nabla F_{m-1}(\mathbf{X}) L\big(y, F_{m-1}(\mathbf{X})\big) = \nabla \hat{y}_{m-1} L\big(y, \hat{y}_{m-1}\big) \qquad (6)$$

LightGBM is a variant of the GBM algorithm that uses two techniques to improve its performance and efficiency: GOSS (Gradient-based One-Side Sampling) and EFB (Exclusive Feature Bundling) [14]. GOSS works by downsampling data instances in order to maintain the accuracy of the information gain estimate. To do this, it keeps instances with large gradients (large errors), while randomly eliminating instances with small gradients. Consequently, it is possible to improve the efficiency of the model by focusing attention on the most important instances. EFB is used to reduce the amount of data needed in a sample, through the combination of resources. Different from GBM, LightGBM uses a histogram-based method that is faster and more efficient in finding features and segmentation boundaries, it discretizes data characteristics into a histogram algorithm, and uses the discretized value as an index to accumulate statistics. After reading the data, the histogram accumulates the necessary statistics and, according to the discrete value of the histogram, finds the optimal split point. In this way, LightGBM can reduce calculation time and achieve solutions closer to ideal. This algorithm is known as the leaf split method, which results in deeper trees and fewer leaves, allowing more complex patterns to be captured in the data, improving accuracy and reducing training time [14].

The performances of the aforementioned algorithms, for our case, are consolidated and demonstrated in section IV.

B. Interpretable Model

With the development of models, the need for interpretability arises, which refers to the ability to understand how predictive methods achieve their results [15]. This topic is considered as a sub-area in machine learning research, and referred to as explanatory artificial intelligence (XAI) [16]. In the literature there are some algorithms developed for this study, such as SHAP (SHapley Additive exPlanations) and LIME [17,18]. In this study, we chose to use LIME considering the need for local interpretability to individually understand each client, without requiring high processing power and high complexity as in SHAP algorithms [19].

To mathematically understand the operationalization of the algorithm, the objective of LIME is: given a predictive model f(x), LIME will look for the contributions of each variable through a local linear approximation given by a new model g(x). For each y predicted in a model f(x), LIME will create a function g(x) given through new linear models applied locally. For this, the algorithm consists of taking a local sample of the observation to be explained and applying perturbations, creating synthetic samples around x that are classified by the f(x) model. The created samples are weighted by a kernel π according to their proximity to x, in which closer samples receive greater weights than more distant ones. Using these samples with their respective weights as well as the values of the original model applied to them, a linear model g(x) is trained to explain the result of sample x. The LIME algorithm seeks to minimize the error $L(f,g,\pi_x)$ given by the difference between the response of the original predictor model f(x) and the local predictor model g(x). Equation 7 describes the behavior of LIME, in which G is the class of linear models. The element Ω(g) is a measure of complexity of the explanation generated by g, in which it guarantees local interpretability, or the loss of locality recognition while maintaining low enough to be interpreted by humans [20].

$$\text{explanation}(x) = \arg\min L(f, g, \pi_x) + \Omega(g) \qquad (7)$$

where: $g \in G$

C. *Unsupervised Learning*

Clustering is an unsupervised learning technique that groups data into clusters based on similarities between them [21]. There are several clustering algorithms, such as K-means, hierarchical clustering and DBSCAN. However, choosing the best algorithm depends on the type of data and the objective of the analysis. In the present study, LIME values represent the contribution of each variable to each customer's churn. As this is continuous, non-categorical information and it is a non-parametric approach, the algorithm chosen was K-means [21].

K-means is a partitional unsupervised learning method that divides data into K distinct groups, where K is previously defined, and its dissimilarity measure is based on the Euclidean distance. After choosing K, the algorithm works in three steps: first, the arbitrary initialization of K centroids; second, assign each sample to its closest centroid; third, recalculate the cluster centroids with the assignments produced in the second step and repeat the second and third steps until there are no significant changes in the centroid values [22].

Mathematically, given a set of N objects $X=\{x_1,x_2,\ldots,x_N\}$ and a set of k vectors $\{x_1,x_2,\ldots,x_k\}$ representing the centroids, the K-means algorithm has the objective function of minimizing the value of sum of errors represented by J, given by equation (8), in order to reduce the distance between elements of the same cluster, indicating the grouping of similar customers, in which μ_{cj} indicates the probability of the object x_j belongs to the cluster x_c, such as it satisfy the normalization condition expressed by equation (9). The premise adopted for this study is that each client could only belong to one cluster, so μ_{cj} becomes a binary value [23].

$$J = \sum_{j=1}^{N} \sum_{c=1}^{k} \mu_{cj} \left\| \mathbf{X}_j - \bar{\mathbf{X}}_c \right\|^2 \qquad (8)$$

$$\sum_{c=1}^{k} \mu_{cj} = 1 \ \forall j \qquad (9)$$

It is important to highlight that the choice of the number of clusters is not trivial and they can be done in several ways. The strategy adopted in this work was based on the Elbow method, which consists of varying the number of clusters and comparing the cluster inertia variation curve, in other words, through intra-cluster variability in relation to the number of clusters, identifying where the variability gain decreases significantly [24]. The results of this method applied to our data are presented in section IV.

III. Methodology

The proposed solution involved the following steps: Development of the machine learning model, Interpretability of the model with LIME and clustering using K-means where the each of them will be described in the next sessions.

A. *Churn Prediction Modeling*

This study was applied to an e-commerce database, in which the non-disclosure of the company's name is essential to guarantee the confidentiality and protection of sensitive data related to customers and operations of the company in question. The base consists of a sample of approximately 405 thousand random customers who carried out purchasing activities in the month in which observations of the predictor variables were collected and in the subsequent month it was observed whether the customer purchased again or not. In total there was a churn rate of 87.6%, that is, approximately 355 thousand customers were marked as churning and 50 thousand made the purchase again, representing 12.4%. The dataset originally has 46 variables, with two identifier variables: customer ID and date and time of last purchase. As for the other variables, 20 are categorical, 2 are Boolean and 22 are numeric variables. Observations that contained missing values were removed from the analysis, representing 0.7% of the total base. As redundancy and irrelevant variables in the data set can harm the predictive performance of the model, variables were selected using the concepts of Information Value (IV) and WOE (weight of evidence) for dimension reduction and correlation analysis to remove redundant variables.

Machine learning classification models were developed and trained. For comparison criteria between models, the number of true positives (TP) and true negatives (TN) were used as inputs for subsequent metrics, that is, the number of correct classifications the model made, and the number of false positives (FP) and false negatives (FN), which indicates the number of incorrect classifications that the model made. Based on this information, the following parameters were used for comparative metrics:

- ROC (Receiver Operating Characteristic) curve and AUC (Area Under Curve): the ROC curve is a graph that shows the relationship between the true positive rate (TPR - True Positive Rate) and the false positive rate (FPR - False Positive Rate) at different classification thresholds. AUC refers to the area below the ROC curve, which evaluates the ability of a classification model to distinguish between positive and negative classes [25].
- Accuracy (identified by Acc) measures the proportion of correct predictions in relation to the total number of predictions made by the model [26]:

$$\text{Acc}=\frac{\text{TP + TN}}{\text{TP + TN + FP + FN}} \qquad (10)$$

- Precision measures the proportion of true positives in relation to the total positives predicted by the model [27]:

$$\text{Precision}=\frac{\text{TP}}{\text{TP + FP}} \qquad (11)$$

- Recall measures the proportion of true positives in relation to the total number of real positives in the database [26]:

$$\text{Recall}=\frac{\text{TP}}{\text{TP + FN}} \qquad (12)$$

- F1-score is the harmonic mean between precision and recall, given by the formula:

$$\text{F1}_{\text{score}}=\frac{\text{2.Recall.Precision}}{\text{Recall + Precision}} \qquad (13)$$

B. *Interpretability and Clustering*

To identify the impact criteria that each variable contributed to each customer's churn decision, providing transparency and understanding about the model's predictions, local explanations of the model were applied using LIME to evaluate the relative importance of each characteristic, its result it is a catalog of marginal probabilities, negative and positive, that each feature contributes to the model diagnosis for that example [10]. With the output generated, the result was applied as input in a K-means clustering algorithm, as a hypothesis was raised that customers prone to canceling their contracts can be separated into distinct clusters according to their motivations (it is important to highlight that only prone customers were grouped). To define the number of clusters, the Elbow Chart was used.

IV. Results and Analysis

After the variable selection stage, 11 of them remained as predictors of the model: gender, age, registration in a loyalty program, time since the last purchase made, number of complaints in the last 6 months, number of accesses made in the last month on the platform, number of coupons applied in the last 6 months, customer registration region (variable associated with shipping and delivery time), number of items returned and refund history and average purchase value. The information values (IV) obtained by each of these variables is shown in Table I.

TABLE I. Information Value

Variable	IV
Number of days since the last purchase	0.440785
Region registered for delivery	0.380952
Number of complaints	0.299288
Average ticket	0.199976
Number of accesses in the last month	0.188311
Registration in the loyalty program	0.136226
Age	0.103497
Number of returns	0.071754
Refund history	0.065319
Number of coupons used	0.038953
Gender	0.038270

With these variables, the prediction models were developed and their performances are demonstrated in Table II. It is noted that the boosting models stand out in terms of AUC, Accuracy and F1-Score in relation to the others, however the LightGBM model when contrasted with the GBM outperforms, in addition to requiring less processing power, achieving a maximum accuracy of 71%, F1-Score of 71% and AUC of 78%.

TABLE II. Model Performance

Modelo	AUC	Acc	Precision	Recall	F1-Score
Light Gradient Boosting	0.783	0.710	0.715	0.708	0.711
Gradient Boosting	0.776	0.705	0.694	0.710	0.702
Random Forest	0.765	0.697	0.701	0.696	0.698
Logistic Regression	0.734	0.681	0.664	0.687	0.675

With the model applied to the customer base, only customers with a propensity to churn were selected to apply LIME and subsequently clustered. Using the Elbow method, it was identified that 5 clusters would be sufficient for the case in question (Figure 1), in which the distribution of the number of customers in each cluster is demonstrated in Figure 2.

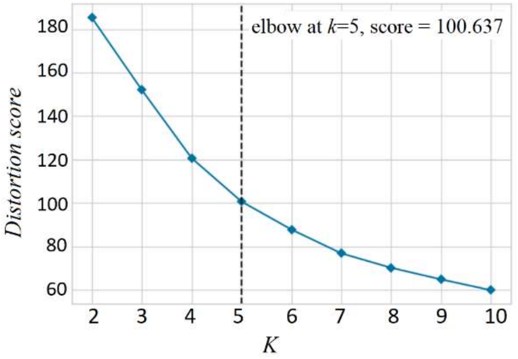

Fig. 1. Elbow Chart

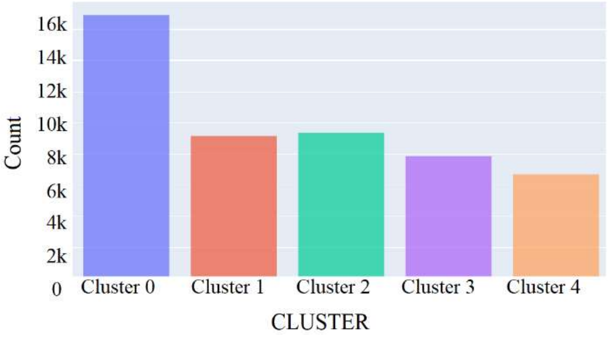

Fig. 2. Distribution of clustered customers

For each cluster, a graph was created that demonstrated the degree of impact of each variable and the relative number of customers (Figure 3). It shows three variables with high impact power: number of days since the last purchase, number of complaints and region registered for delivery, as well as the low impact of gender and age variables.

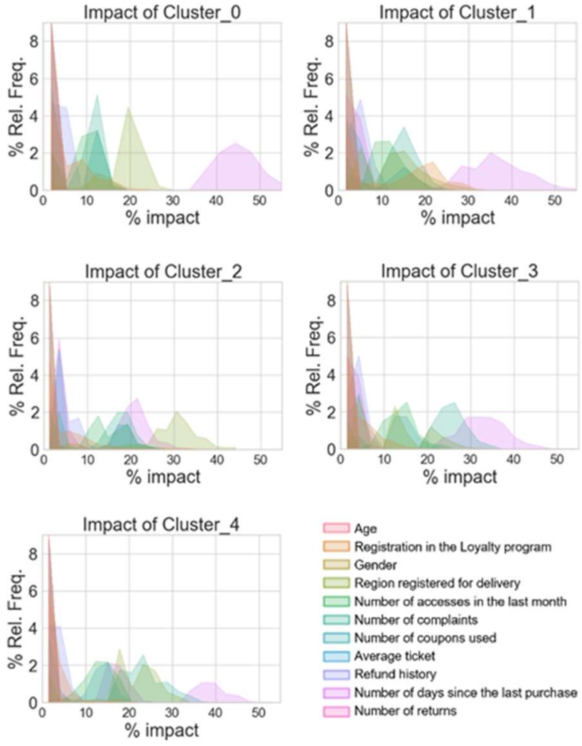

Fig. 3. Distribution of impact in each Cluster

It is noted that cluster 0 stands out for the extremely high impact of the number of days since the last purchase, exceeding 50% in the influence of the propensity to churn, demonstrating a customer profile with high downtime, in addition to having impacts on the variables number of complaints and registered region, indicating the possibility of being a cluster of customers who had bad experiences and did not purchase again. As for cluster 1, in addition to the high impact of downtime, the variable loyalty program subscriber stands out, indicating the loss of interest in the benefits of the program and the history of problems in the number of complaints. When observing Cluster 2, we notice the inversion of the degree of impact between the variables time since the last purchase and the registered region, indicating that this group of customers has a strong impact due to issues related to shipping, whether in price or delivery time, which could be a group of customers with a region remote from the logistics distribution center and pointing out a possibility for a strategy focused on this topic. Finally, in Cluster 3, there is an increase in the impact of the variable number of complaints, while in Cluster 4 it indicates a profile of dissatisfied customers in general. In summary, there are clusters:

- Cluster 0 - Inactive Customer: customer with a long period of inactivity, indicating a bad experience in the past and has not purchased again.
- Cluster 1 - Customer Dissatisfied with the loyalty program: this customer archetype indicates a loss of interest in the loyalty program, indicating that the advantages offered are not meeting their expectations.
- Cluster 2 - Customer with Shipping Problems: customer impacted mainly by shipping, whether in terms of price or delivery time.
- Cluster 3 - Customer Dissatisfied with the Product: this archetype presents an increase in the impact of the variable number of complaints, indicating that there are problems with the products.
- Cluster 4 - General Dissatisfied Customer: this customer archetype indicates a profile of general dissatisfaction with the company. It is important to investigate the reasons for this dissatisfaction and take steps to improve the customer experience.

Regarding the practical implications of the results, the identified customer clusters can be used to develop more targeted retention strategies. For example, customers in Cluster 0 can be approached with personalized re-engagement campaigns, offering exclusive discounts or encouraging product reviews to regain their trust. For Cluster 1, companies can reevaluate and adjust their loyalty offers, creating more attractive incentives that are aligned with customer expectations, i.e., offering a loyalty program with personalized rewards based on customer purchase history and preferences. As for Cluster 2, it is suggested that logistics partnerships be negotiated to reduce shipping costs and improve delivery times, prioritizing regions with the highest concentration of customers in this cluster, or that faster and more economical shipping options be introduced for this audience. For customers in Cluster 3, it is crucial to conduct a detailed analysis of the products with the highest number of complaints and implement improvements in the quality and control of these items. Finally, for Cluster 4, an in-depth investigation of the causes of dissatisfaction is essential, which may include improvements in customer service, the return and refund process, and communication with customers. For this, data mining techniques could help in collecting valuable insights if applied in this cluster.

In order to test the robustness and ensure the generalizability and applicability of the proposed methodology in different e-commerce environments, three additional data sets from different months of the same company were used, with a period of 2 months between each base. The aim was to ensure customer variation, as well as different consumption and behavior profiles. In each of these sets, the steps of the proposed methodology were repeated, from the construction of the predictive churn model, application of the LIME interpretability technique, to K-means clustering. The results showed that, even with data variability, the model maintained consistent performance in terms of accuracy and precision without significant variations, eliminating the need for adjustments in the hyperparameters. This consistency indicates the stability of the model, which can be effectively applied in different contexts and periods. Regarding the comparison with other churn prediction methods, such as Logistic Regression, for example, the analysis revealed that although its interpretability is easier, its performance was inferior to that of the Machine Learning model chosen in this work. Therefore, the advantage of using LIME is that it does not limit the prediction technique to be chosen, allowing the use of more robust models. Regarding the clustering technique used, when compared with other types, K-means stood out for its simplicity and effectiveness in forming cohesive and actionable groups, without requiring the application of other techniques, such as DBSCAN, which has difficulties in dealing with the high dimensionality of the data [28], or Hierarchical Clustering, which is less scalable due to its algorithm [29].

The results demonstrated that the proposed approach provided more interpretability and coherent customer segmentation, facilitating the identification of common motivations and the implementation of more targeted retention strategies. Therefore, compared to other methods that use only predictive models for churn [2][3][4][5] or that use clustering alone to create customer segmentation [28][29], the present approach stands out for its ability to provide actionable insights and a deep understanding of the factors that influence customer churn by combining these machine learning techniques.

V. Conclusion

The initial objective of the work was achieved, making it possible to generally identify prone customers through the machine learning model and the subsequent construction of groups of common motivations in favor of K-means applied to the model's interpretability. In contrast to most churn prediction models, the interpretability of the model via LIME makes it possible to add to commonly used predictive model solutions the traceability of potential problems that may be motivating customers to churn, enabling implementation by companies, the creation of a strategy for proactive and personalized intervention for each customer at risk of churn. Another difference is the way this research uses K-means clustering, unlike conventional approaches that apply the algorithm to the customers' raw variables, the application to the LIME output allowed us to identify groups with the same degrees of impact on each variable of their motivation to churn.

As a challenge faced, the selected data set lacks more variables with greater predictive potential so that the resulting centroids have discrete strategic implications, since the models metrics could be improved and the differentiation and discretion of each cluster as well. Therefore, future studies can utilize a more robust dataset to validate and employ the K-means methodology to perform a more complete analysis of the customers represented by individual centroids. This makes it worth the challenge of applying it to business bases that rely on Big Data, in which there is greater complexity in working on the individual motive of each client, given the greater possibility of variables and the heterogeneity of profiles. Likewise, the methodology can be applied on other fronts besides retention, such as sales or product upgrades, for example.

Even given the limitation of immediate practical application, the study offers valuable insights that can guide future strategies, contribute to the generalization of findings in other industries or sectors and generate implications for business practice. In addition, to providing a methodological model that can serve as a basis for subsequent studies and practical experiments.

REFERENCES

[1] J. Shobana, Ch. Gangadhar, R. K. Arora, P. N. Renjith, J. Bamini, and Y. D. Chincholkar, “E-commerce customer churn prevention using machine learning-based business intelligence strategy,”, Measurement: Sensors, vol. 27, p. 100728, ISSN 2665-9174, 2023, doi: 10.1016/j.measen.2023.100728

[2] F. A. Mohamed and A. K. Al-Khalifa, "A Review of Machine Learning Methods For Predicting Churn in the Telecom Sector," 2023 International Conference On Cyber Management And Engineering (CyMaEn), Bangkok, Thailand, 2023, pp. 164-170, doi: 10.1109/CyMaEn57228.2023.10051108.

[3] A. Khattak, Z. Mehak, H. Ahmad, M. U. Asghar, M. Z. Asghar, and A. Khan, ‘‘Customer churn prediction using composite deep learning technique,’’ Sci. Rep., vol. 13, no. 1, Oct. 2023, Art. no. 17294, doi: 10.1038/s41598-023-44396-w.

[4] P. Berger and M. Kompan, "User Modeling for Churn Prediction in E-Commerce," in IEEE Intelligent Systems, vol. 34, no. 2, pp. 44-52, March-April 2019, doi: 10.1109/MIS.2019.2895788.

[5] X. Xiahou and Y. Harada, “B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM.”, Journal of Theoretical and Applied Electronic Commerce Research, vol. 2, no. 2, pp. 458–475, 2022, doi: 10.3390/jtaer17020024.

[6] I. AlShourbaji, N. Helian, Y. Sun, A. G. Hussien, L. Abualigah, and B. Elnaim, ‘‘An efficient churn prediction model using gradient boosting machine and metaheuristic optimization,’’ Sci. Rep., vol. 13, no. 1, p. 14441, Sep. 2023, doi: 10.1038/s41598-023-41093-6.

[7] P. P. Singh, F. I. Anik, R. Senapati, A. Sinha, N. Sakib, and E. Hossain, “Investigating customer churn in banking: A machine learning approach and visualization app for data science and management,”, Data Science and Management, vol. 7, no. 1, pp. 7–16, ISSN 2666-7649, 2024, doi: 10.1016/j.dsm.2023.09.002.

[8] Z. Wu, L. Jing, B. Wu, and L. Jin, ‘‘A PCA-AdaBoost model for e-commerce customer churn prediction,’’ Ann. Oper. Res., vol. 2, pp. 1–18, 2022, doi: 10.1007/s10479-022-04526-5.

[9] P. Kotler and K. L. Keller, “Administração de Marketing”, 3th ed., São Paulo, Pearson Education do Brasil, 2019.

[10] C. Sun, J. Sharma and M. Maiti, "Leveraging Machine Learning and Model-Agnostic Explanations to Understand Automated Diagnosis of Cardiovascular Disease," 2022 4th International Conference on Biomedical Engineering (IBIOMED), Yogyakarta, Indonesia, 2022, pp. 36-41, doi: 10.1007/s10791-011-9176-2.

[11] R. C. Sonawane and H. D. Patil, "Clustering Techniques and Research Challenages in Machine Learning," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2020, pp. 290-293, doi: 10.1109/ICCMC48092.2020.ICCMC-00054.

[12] K. Makkar, P. Kumar, M. Poriye and S. Aggarwal, "A Comparative Study of Supervised and Unsupervised Machine Learning Algorithms on Consumer Reviews," 2022 IEEE World Conference on Applied Intelligence and Computing (AIC), Sonbhadra, India, 2022, pp. 598-603, doi: 10.1109/AIC55036.2022.9848880.

[13] J. H Friedman, "Greedy function approximation: A gradient boosting machine", Ann. Statist., vol. 29, pp. 1189-1232, 2001.

[14] G.L. Ke, Q. Meng, T. Finley, T. F. Wang, W. Chen, W.D. Ma, et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree", Advances in Neural Information Processing Systems, pp. 3149-3157, 2017.

[15] N. Barr Kumarakulasinghe, T. Blomberg, J. Liu, A. Saraiva Leao and P. Papapetrou, "Evaluating Local Interpretable Model-Agnostic Explanations on Clinical Machine Learning Classification Models," 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), Rochester, MN, USA, 2020, pp. 7-12, doi: 10.1109/CBMS49503.2020.00009.

[16] Explainable artificial intelligence (XAI), Darpa, Aug. 2016, [online] Available: https://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf.

[17] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions", Proc. Adv. Neural Inf. Process. Syst., vol. 30, pp. 1-10, 2017.

[18] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, et al., "From local explanations to global understanding with explainable AI for trees", Nature Machine Intelligence, vol. 2, no. 1, pp. 56-67, 2020, doi: 10.1038/s42256-019-0138-9.

[19] A. Bora, R. Sah, A. Singh, D. Sharma and R. K. Ranjan, "Interpretation of machine learning models using XAI - A study on health insurance dataset," 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2022, pp. 1-6, doi: 10.1109/ICRITO56286.2022.9964649.

[20] M. T. Ribeiro, S. Singh and C. Guestrin, "Why should i trust you?", Explaining the predictions of any classifier” In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135-1144, August 2016.

[21] K. P. Sinaga and M. -S. Yang, "Unsupervised K-Means Clustering Algorithm," in IEEE Access, vol. 8, pp. 80716-80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[22] S. Shankar, B. D. Sarkar, S. Sabitha and D. Mehrotra, "Performance analysis of student learning metric using K-mean clustering approach K-mean cluster," 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence), Noida, India, 2016, pp. 341-345, doi: 10.1109/CONFLUENCE.2016.7508140.

[23] J. MacQueen, ‘‘Some methods for classification and analysis of multivariate observations,’’ in Proc. 5th Berkeley Symp. Math. Statist. Probab., Oakland, CA, USA, 1967, pp. 281–297.

[24] D. Marutho, S. Hendra Handaka, E. Wijaya and Muljono, "The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News", 2018 International Seminar on Application for Technology of Information and Communication, pp. 533-538, Sept. 2018.

[25] J. A. Hanley and B. J. McNeil, “The meaning and use of the area under a receiver operating characteristic (ROC) curve,” Radiology, vol. 143, no. 1, pp. 29–36, 1982. doi: 10.1148/radiology.143.1.7063747.

[26] D. M. W. Powers, “Evaluation: From precision, recall and F-factor to ROC, informedness, markedness & correlation,” Flinders University, Bedford Park, SA, Australia Tech. Rep., 2007, doi: 10.48550/arXiv.2010.16061.

[27] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” Information Processing & Management., vol. 45, no. 4, pp. 427–437, ISSN 0306-4573, 2009, doi: 10.1016/j.ipm.2009.03.002.

[28] A. S. M. S. Hossain, "Customer segmentation using centroid based and density based clustering algorithms," 2017 3rd International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, 2017, pp. 1-6, doi: 10.1109/EICT.2017.8275249.

[29] A. Afzal et al., "Customer Segmentation Using Hierarchical Clustering," 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), Pune, India, 2024, pp. 1-6, doi: 10.1109/I2CT61223.2024.10543349.