An Adaptive Clustering-Based Customer Loyalty Score for Modeling of Customers' Behavior

Marcello Eduardo Mônaco
Tecnologia em Bancos de Dados
Faculdade Impacta de Tecnologia (FIT)
São Paulo – SP – Brazil
marcello.monaco@aluno.faculdadeimpacta.com.br

Alex Torquato Souza Carneiro
Tecnologia em Bancos de Dados
Faculdade Impacta de Tecnologia (FIT)
São Paulo – SP – Brazil
alex.carneiro@faculdadeimpacta.com.br

Alan Felipe R. Reggi de Oliveira
Tecnologia em Bancos de Dados
Faculdade Impacta de Tecnologia (FIT)
São Paulo – SP – Brazil
alan.reggi@aluno.faculdadeimpacta.com.br

Gabriely Fagundes O. Martins
Tecnologia em Bancos de Dados
Faculdade Impacta de Tecnologia (FIT)
São Paulo – SP – Brazil
gabriely.martins@aluno.faculdadeimpacta.com.br

Nicolly Nascimento A. Trajano
Tecnologia em Bancos de Dados
Faculdade Impacta de Tecnologia (FIT)
São Paulo – SP – Brazil
nicolly.araujo@aluno.faculdadeimpacta.com.br

Abstract—A customer loyalty score indicates the strength of the relationship between a customer and a seller, e.g., branches, makers, and stores. Conventional loyalty scores are calculated through linear mathematical models, for example, a linear equation that returns the score as a weighted sum of the Recency, Frequency, and Money values obtained from the RFM approach. Linear models present several restrictions for customer segmentation because they associate very close scores with customers in distinct patterns of Recency, Frequency, and Money. In this work, we introduce a novel customer loyalty model based on clustering that calculates the loyalty of a customer according to his/her adherence to each cluster. The obtained results suggest our proposal yields a discriminative score that "summarizes" the strength of the relationship between customers and sellers.

Index Terms—customer behavior, customer profiles, clustering, mathematical analysis

I. Introduction

In today's hyper-competitive landscape, corporate growth depends not only on the quality of the products and services offered. The ability to take data-driven decisions has become an even more critical determinant of success and in some cases surviving. Organizations that transform data into actionable insights—leveraging predictive analytics, sophisticated segmentation, and behavioral modeling can achieve a more granular understanding of their customers, anticipate needs proactively, and cultivate a competitive advantage that rivals find difficult to match [1], [2].

Calculating a customer loyalty score gives to sellers a single actionable metric that captures how committed (and profitable) each customer is, and also likely will be in the future. This kind of score helps identify high-value customers that generate most of the revenue, allowing significant optimizations in marketing and retention resources. A loyalty score also serves as an early warning system for churn, enabling proactive engagement before valuable customers are lost. By quantifying customer engagement, it supports personalized experiences at scale, such as dynamic offers, product recommendations, or loyalty program tiers, customized to each shopper's value [3]. In addition, it improves the accuracy of the prediction, helps measure the *return on investment* (ROI) of retention initiatives, and identifies brand advocates who can fuel referrals and user-generated content. Ultimately, the loyalty score becomes a shared metric across teams, aligning product, marketing, and customer experience efforts around long-term customer value and business growth.

Usually, loyalty score are based on Recency, Frequency, and Money (RFM) values and employ a kind of static (mostly linear) model to produce the loyalty score [4]. These models do not consider that customers' behavior evolves due to seasonality, marketing actions, product changes, external events, or life changes. Due to the dynamics in customers' behavior, an adaptive model might be more useful to produce the loyalty score.

At this work we introduce a novel method for calculating customers' loyalty scores supported by customers clustering. In order to get useful scores for loyalty, ensuring that each cluster of customers represents truly distinct profiles, we looked for obtaining minor overlap of scores across different clusters. Finally, the proposed method produces loyalty scores that adapt themselves to the customers' patterns of RFM values, and also makes possible to update the scores when the stakeholders consider necessary.

Next sections of this work are organized as follows: Section II compiles relevant works related to application of clustering techniques for customers grouping, loyalty scores and the employed evaluation dataset, Section III describes the proposed method and how each step works in a coherent pipeline, Section IV contains the evaluation methodology and metrics, Section V presents the obtained results, Section VI discusses the obtained results highlighting advantages and disadvantages of the proposed method, and Section VII presents the final conclusions of this work and also proposes some future works.

II. RELATED WORKS

Customer loyalty scoring has evolved significantly beyond traditional RFM (Recency, Frequency, Monetary) models, in retail this kind of score is largely applied to several aspects, such as optimizations in marketing and retention resources, and ROI measurement. While RFM remains foundational due to its simplicity and interpretability, modern approaches now integrate machine learning to uncover non-linear and complex relationships between purchase behaviors and loyalty. However, typical loyalty scores are based on linear models [5]–[7].

Otherwise clustering techniques, like K-Means, have been employed and presented promising results in a number of works [8]–[10]. Clustering models are mostly employed for customers' segmentation, that is a useful task for marketing campaigns. Customer segmentation and loyalty scoring are both essential tools in customer analytics, but they serve distinct purposes and operate differently. Customer segmentation involves grouping customers based on shared characteristics or behaviors, such as demographics, purchasing habits, or RFM values. The goal of segmentation is to identify patterns among different customer types to tailor marketing strategies, product offerings, or communication. For example, segmentation can help distinguish between price-sensitive buyers and premium shoppers or between new and long-term customers. It may be seen as a classification task that helps businesses understand who their customers are and how to engage with them more effectively [1].

In contrast, loyalty scoring is a predictive or evaluative approach that assigns a numerical value or grade to each customer, reflecting their likelihood to remain engaged, repeat purchases, or contribute long-term value. While segmentation places customers into discrete groups, loyalty scoring ranks them on a continuum—enabling finer-grained prioritization. Loyalty scores can be calculated using a range of features, including RFM data, purchase trends, and engagement history, often with weighting or machine learning techniques [4]. This allows companies to assess not just "who" their customers are, but "how committed or valuable" each one is likely and an expectation about how he/she will be in the future, which is critical for retention strategies, loyalty programs, and resource allocation. However, the use of clustering models as support for loyalty scoring has not been explored in existing works, although it might be insightful and provide high adaptive scores.

Recent studies, such as [11], have evaluated hybrid models combining machine learning methods and LightGBM (RF-LightGBM) to improve repurchase prediction by addressing data imbalance, hyperparameter tuning, and feature complexity in large e-commerce dataset. These models significantly outperform classical methods by capturing behavioral nuances that static RFM segments miss. This kind of approach may also provide promising adaptive loyalty scoring, but the existing works have explored other tasks than loyalty scoring, e.g. repurchase prediction and customer retention.

Due the relevance of loyalty scoring and the potential advantages of an adaptive scoring method, we propose a clustering-based scoring able to adapt customers' scores according to the patterns found into the training dataset. The proposed score should be both highly discriminant for customers' segmentation and also provides profitability insights.

To evaluate the proposed method, we used a well known public dataset available at UC Irvine Machine Learning Repository *UCI Online Retail II* [12], which has been used in several studies, such as [13]–[16]. This dataset contains two years of invoice-level sales records from an online retailer from United Kingdom. Columns used in this work were as follows: Invoice, Customer ID, Quantity, Price and InvoiceDate.

The previous works that investigated customers behavior using the Online Retail II dataset aimed to apply segmentation techniques to create groups of customers with similar RFM patterns and then conduct analysis in groups level. Our proposed method is able to support customers segmentation with some precision and also provide a global vision of the customers and their profitability for the seller.

III. PROPOSED ADAPTIVE CLUSTERING-BASED CUSTOMER LOYALTY SCORE

The foundations of the proposed method are in the RFM concepts that stands for Recency, Frequency, and Monetary. They mean three dimensions used to evaluate customer engagement and value. **Recency** refers to how recently a customer made their last purchase, indicating their current interest or connection to the brand. Customers who purchased more recently are typically more likely to respond to marketing efforts and return to buy again. **Frequency** measures how often a customer has made purchases within a specific period. A higher frequency suggests greater engagement and loyalty, signaling that the customer has formed a habit or preference for the brand's products or services. **Monetary** refers to the total amount a customer has spent over a defined timeframe, helping businesses identify high-value customers who contribute more to revenue.

Together, these three metrics offer a comprehensive view of customer behavior. The RFM values are calculated per customer (c) as follows:

$$
\begin{aligned}
\text{Recency}_c &= (\text{LAST.DATE} - \max\{\text{InvoiceDate}\}), \\
\text{Frequency}_c &= |\{\text{Invoice}_{(c,t)}\}|, \\
\text{Monetary}_c &= \sum_t \text{TotalNF}_{(c,t)},
\end{aligned}
\tag{1}
$$

where $LAST.DATE$ reflects the last date of valid transactions in the dataset, (t) indexes transactions' time and $|\{\text{Invoice}_{(c,t)}\}|$ means the number of transactions within the dataset's time window.

Given the differences in scale of the RFM values among the distinct customers in a dataset, we apply the z-score normalization:

$$
\tilde{R} = \frac{R - \mu_R}{\sigma_R}, \ \tilde{F} = \frac{F - \mu_F}{\sigma_F} \ and \ \tilde{M} = \frac{M - \mu_M}{\sigma_M}.
\tag{2}
$$

Where μ and σ are the mean and standard deviation of each set of values, respectively.

Due to the frequent outliers, we needed to apply a transformation in $\tilde{R}$, $\tilde{F}$ and $\tilde{M}$ values. The hyperbolic tangent ($\tanh$), defined as

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \tag{3}$$

was chosen because it is monotonically crescent sigmoid-shaped function, asymptotically limited between $(-1, 1)$ and keeps the 0 (zero) value as the original position.

The hyperbolic tangent transformation was applied to $r_h = \tanh(\tilde{R})$, $f_h = \tanh(\tilde{F})$ and $m_h = \tanh(\tilde{M})$, yielding values limited between $(-1, 1)$. The hyperbolic tangent keeps the original 0 (zero) value and compresses the values while maintaining their sort order.

The transformed dataset is represented by a table with columns $\{r_h, f_h, m_h\}$, where all values are limited in $(-1, 1)$ and the 0 value represents the mean of original R, F and M values. This dataset feeds a K-Means model that calculates K centroids in the $\{r_h, f_h, m_h\}$ space. At this point we propose the heuristic algorithm for sorting the clusters' centroids described in Algorithm 1.

Algorithm 1: Heuristic clusters' centroids sorting.

```
/* centroids is a list with the K-Means
   centroids.                          */
```
Data: $centroids$
Result: $list$
$list \leftarrow$ empty list
for $centroid\ in\ centroids$ **do**
  $r_h, f_h, m_h \leftarrow centroid$
  if $f_h > 0$ **then**
    | append f_h to $list$
  else
    if $m_h > 0$ **then**
      | append m_h to $list$
    else
      if $r_h > 0$ **then**
        | append r_h to $list$
      else
        | append f_h to $list$; // uses f_h if $r_h, f_h, m_h \leq 0$.
      end
    end
  end
end
$sort(centroids)$ according to $list$

By our heuristics, **Frequency** is the most relevant value for the loyalty scoring, followed by **Monetary**, and **Recency** is the least relevant of all. After sorting the centroids, we define the loyalty scores associated to first and last centroids, respectively: 0.1 and 0.9. By doing this attribution of values, we assure all loyalty scores calculated will be within the range

$(0, 1)$. The loyalty scores associated to the remaining centroids are calculated as described in Algorithm 2.

Algorithm 2: Calculus of the loyalty scores for the remaining centroids.

```
/* centroids is a list with the K-Means
   centroids sorted by the heuristic
   algorithm.                          */
```
Data: $centroids$
Result: $loyalties$
$total\_distance \leftarrow 0$
$loyalties \leftarrow$ empty list
append 0.1 to $loyalties$
```
/* Loop from 2 to K for a set of K
   clusters.                           */
```
for $i\ in\ 2 .. K$ **do**
  $aux \leftarrow distance(centroids[i], centroids[i-1])$;
  // distance between consecutive centroids.
  $total\_distance \leftarrow total\_distance + aux$
end
$\delta \leftarrow 0.8$;   // difference from first and last scores.
$accumulator \leftarrow 0.1$
for $i\ in\ 2 .. K$ **do**
  $aux \leftarrow distance(centroids[i], centroids[i-1])$;
  // distance between consecutive centroids.
  $score \leftarrow \frac{\delta \times aux}{total\_distance} + accumulator$
  append $score$ to $loyalties$
  $accumulator \leftarrow score$
end

Finally, a weighted sum of the loyalty scores is calculated for defining the loyalty of each customer as described in Algorithm 3. We chose a Gaussian weighting for the distances to improve the contribution from the closest centroid of each customer, while far centroids provide minor contributions. In the Gaussian weighting, we also use the standard deviation (σ) referent to each centroid, usually referred as cluster **inertia**.

The inertia is a measure of how internally coherent the clusters are, calculated as

$$Inertia = \sum_{c\ \in\ 1..N} \left(distance(RFM_c, centroid_c)\right)^2, \tag{4}$$

where c represents the customer in a set of N customers, and RFM_c and $centroid_c$ mean the RFM values and closest centroid to the customer c respectively. By this way inertia means the sum of squared distances between each data point and the centroid of the cluster to which it belongs.

The inertia reflects the compactness of the clusters: lower inertia indicates that data points are closer to their assigned cluster centers, suggesting tighter and more distinct groupings. Inertia is commonly used as a metric to evaluate the quality of a clustering solution, and it plays a key role in methods

like the "elbow method", which helps determine the optimal number of clusters by identifying the point where adding more clusters no longer significantly reduces inertia.

Algorithm 3: Calculus of the loyalty scores for each customer.

```
/* loyalties is a list with the
   loyalties calculated for the
   centroids.                    */
```
Data: $customers, centroids, loyalties$
Result: $loyalty\_scores$
$weight(d, \sigma) : \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{d^2}{2\sigma^2}}$
$loyalty\_scores \leftarrow$ empty list
for c in $customers$ **do**
    $weight\_sum \leftarrow 0$
    $customer\_loyalty \leftarrow 0$
    for $\{centroid, loyalty\}$ in $\{centroids, loyalties\}$
    do
        $d \leftarrow distance(customer, centroid)$
        $\sigma \leftarrow \sigma_{centroid}$
        $aux \leftarrow weight(d, \sigma)$
        $customer\_loyalty \leftarrow$
         $customer\_loyalty + aux * loyalty$
        $weight\_sum \leftarrow weight\_sum + aux$
        append $\frac{customer\_loyalty}{weight\_sum}$ to $loyalty\_scores$
    end
end

Fig. 1. Elbow curve based on inertia values.

Fig. 2. 3D scatterplot of data with labels from clustering.

IV. EVALUATION

The first processing step is to remove incomplete rows from the dataset for the analysis. Only transactions with positive (> 0) values for quantity and/or price are used in the analysis. A total value per invoice was created, multiplying invoice's Quantity by Price.

Values for RFM metrics were calculated for each customer of the dataset. After calculation of raw RFM data for each customer, Z-score normalization and Hyperbolic Tangent function were applied over data before its usage in clustering. Due to the need of using centroids in our method, K Means was chosen for data clustering. We found the ideal number of clusters is 5 (five) using the Elbow criteria, as presented in Figure 1. The obtained clusters are illustrated in Figure 2, where the clusters are represented by distinct colors.

The customers' loyalty scores were calculated according to the proposed method (described in Section III). We analyzed the scores searching overlapping patterns and comparing the obtained scores against the loyalty scores presented by Aggarwal and Yadav [4]. The achieved results are presented in next section.

V. RESULTS

The obtained loyalty scores for the customers are presented in the histograms of the Figures 3 and 4. These histograms show the concentrati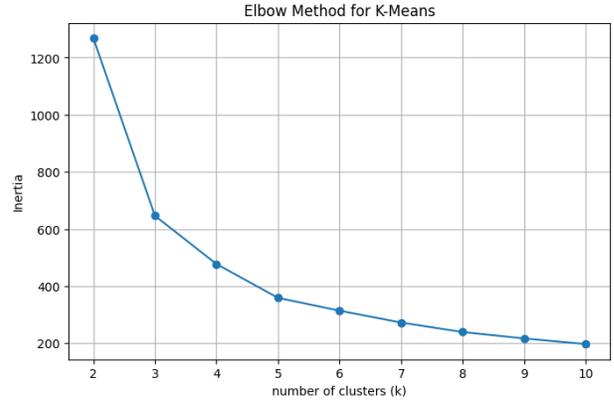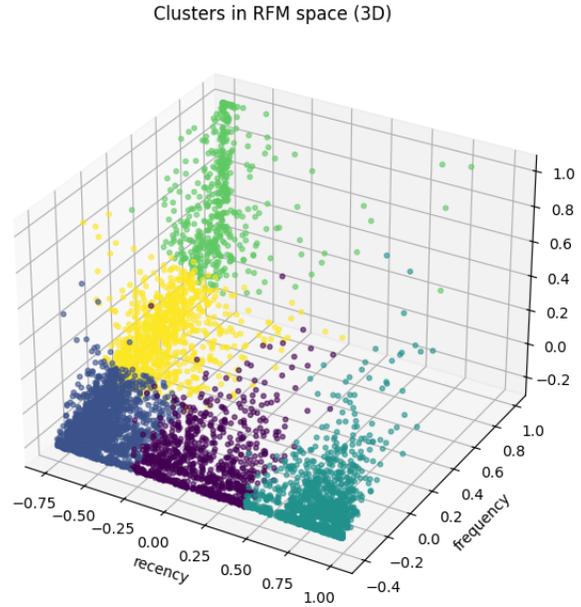on of customers with similar scores into the clusters, where the overlapping and long tail effects might be easily observed.

Histograms present frequency distribution of the customer-level loyalty score, obtained for each of the five K-Means clusters. Each cluster exhibits a single narrow peak whose modal value increases monotonically from cluster 0 (0.10) to cluster 4 (0.90). The tight concentration of observations around their respective modes and the negligible tails beyond the lower bound of the neighboring cluster indicate that the Gaussian distance re-weighting preserved the centroid ordering derived from the R/F/M criteria while minimizing within-cluster variance. The absence of large overlaps among adjacent histograms suggests that the loyalty score provides a graded yet well-separated segmentation of the customer base - an important prerequisite for differentiation in marketing strategies.

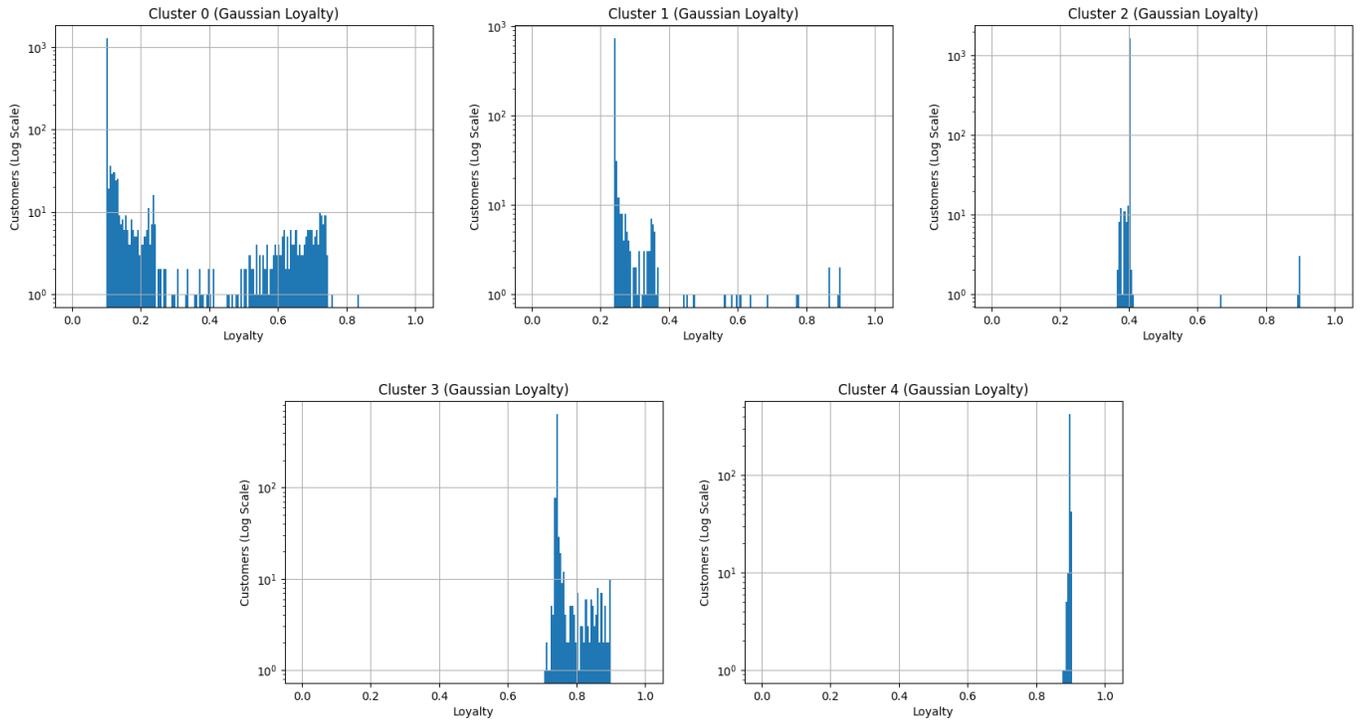In addition to the qualitative analysis, we calculated, for

Fig. 3. Histograms of Gaussian Loyalty Scores by cluster.

each cluster i, the 85th percentile (p_{85}) of its internal loyalty distribution and counted the customers belonging to any other cluster whose loyalty fell within the interval $[\min(i), p_{85}(i)]$. Across all pairwise comparisons, the cross-membership rate remained below 3% (Table I), implying that at least 97% of customers in the densest region of a cluster are unique to that segment. This quantitative criterion, grounded in the cluster-specific σ (standard deviation) bandwidth, supports the claim that the proposed method produces a loyalty metric that is simultaneously continuous at the individual level and sufficiently discriminative for strategic segmentation. Also, same overlap analysis was conducted based on loyalty scores obtained from RFM linear model, generating data presented in Table II that we consider a baseline for comparison.

TABLE I
CROSS-CLUSTER OVERLAP INSIDE EACH CLUSTER'S $[\text{MIN}, P_{85}]$
LOYALTY RANGE - PROPOSED CLUSTERING-BASED SCORING METHOD.

Cluster	Overlapping customers (abs.)	Share of total (%)
0	0	0.00
1	3	0.05
2	10	0.17
3	50	0.87
4	26	0.45
Total	**89**	**1.54**

The baseline results obtained with the linear scoring method were not focused on cluster discrimination, but on customer behavior modeling. Because of that the amount of overlapping samples is much expressive as presented in Table II.

To quantify the degree of overlap between the loyalty

TABLE II
CROSS-CLUSTER OVERLAP INSIDE EACH CLUSTER'S $[\text{MIN}, P_{85}]$
LOYALTY RANGE - LINEAR RFM PROPOSED BY
AGGARWAL AND YADAV [4].

Cluster	Overlapping customers (abs.)	Share of total (%)
0	1928	33.43
1	2917	50.57
2	1308	22.68
3	1004	17.41
4	175	3.03
Total	**7332**	**127.11**

distributions of different clusters, we calculated the pairwise intersection-over-union (IoU) scores across all five clusters, for both linear and gaussian loyalty scores. Each loyalty distribution was previously discretized into a histogram, and the comparison between pairs of histograms was performed using the ratio between the sum of their minimum values (intersection) and the sum of their maximum values (union) across all bins. This metric, commonly used in classification and image segmentation tasks, was adapted here to represent the normalized degree of overlap between customer loyalty profiles. Lower IoU values indicate well-separated distributions, whereas higher values suggest greater similarity and potential overlap between clusters. Results are presented in Figure 5.

Finally, a one-way ANOVA was conducted to determine the effect of using the proposed cluster-based method on the loyalty scoring through the minimum over maximum scores. The results indicate a significant effect, $F(1, 9.01) =$
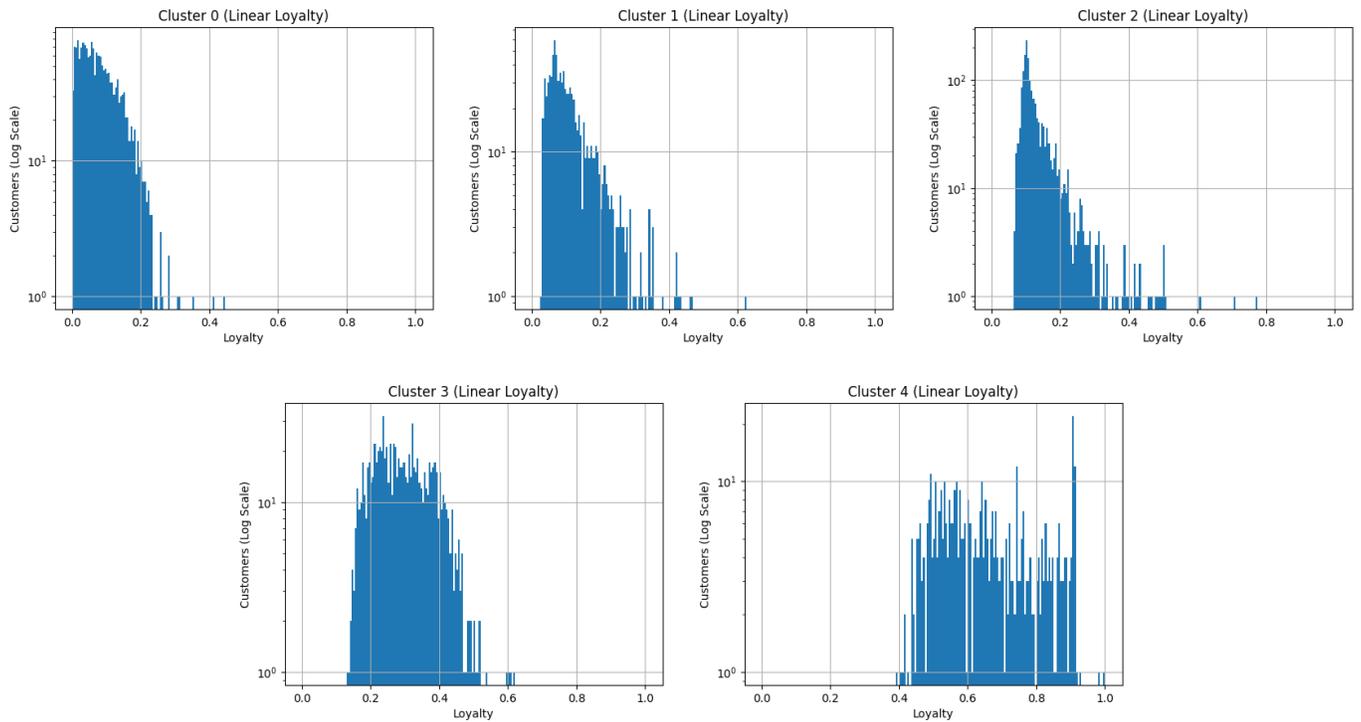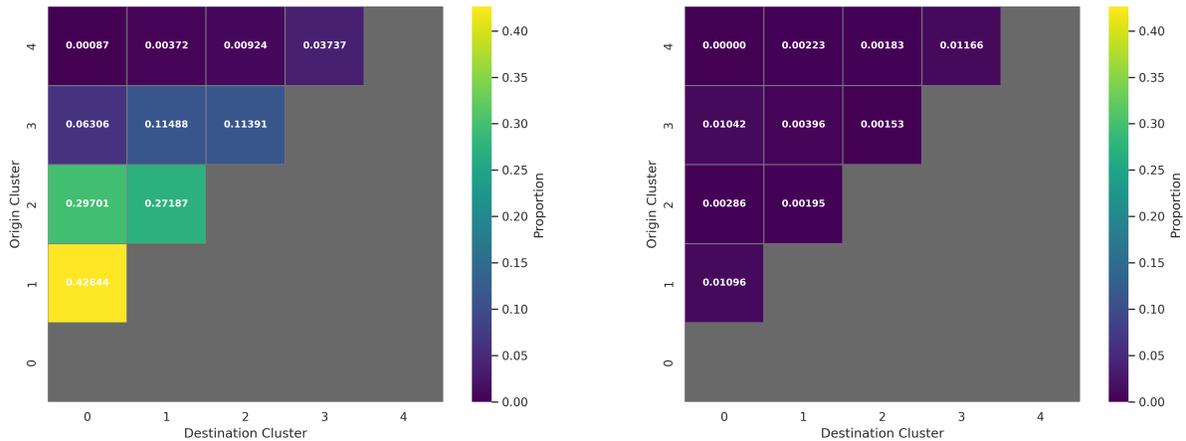
Fig. 4. Histograms of Linear Loyalty Scores by cluster.

Fig. 5. IoU calculation for Linear (left matrix) and Gaussian (right matrix) Loyalty Scores

$7.62, p < 0.1$.

A in-depth analysis of these results is conducted in next section.

VI. DISCUSSIONS

The presented results highlight the discrimination potential of the loyalty scoring method proposed at this work in terms of discrimination and adaptability. However, some limitations might be observed, e.g. the large range of loyalty scores presented by the customers in cluster-0, what happened due to the proximity between these customers to other cluster's centroid with high loyalty score.

In a real scenario, it might be reasonable to perform a second level of clustering for grouping the customers from cluster-0 to avoid this large range of scores. Other alternatives to address large range of loyalty scores are: to increase the number of clusters to likely yield fine-grained scores that will reduce this effect, and investigate distance metrics that might avoid the observed cross clusters influence.

Although the proposed method might require some improvements for providing better results to specific tasks, e.g., for segmentation of customers, the adaptability potential seems useful for uncovering patterns from distinct sets of customers in virtually any industry. Sellers will be able to periodically

update loyalty scores of their customers and define selling actions according to the updated customers' profiles.

About the comparison between our proposed scoring method and the linear scoring method proposed by Aggarwal and Yadav [4], each method has distinct objectives and attend specific demands. While our proposal focuses on discrimination between clusters and adaptability, the linear method models the customers behavior in a static and global perspective that is useful for defining cross-industry standards.

VII. CONCLUSIONS

At this work, we have proposed a clustering-based loyalty scoring supported by RFM model and clustering techniques. Our proposal has been evaluated on a retail public dataset. The proposed method may be applied in an adaptive perspective, reflecting changes in customer behavior during the time. Based on these scores, marketing professionals can take business decisions based on data, addressing specific strategies in order to improve sales and profitability. Based on the results obtained, some points could be highlighted. First of all, it is important to remark that, despite effectiveness of the heuristic adopted at this work, alternative optimization strategies could also be investigated to discover cluster structure or fine-tune loyalty weights.

On the other hand, the proposed method is intrinsically adaptive: by recalculating centroids and Gaussian bandwidths as fresh RFM data arrive, it can operate in an online mode that continually captures shifts in customer behavior that means a very useful information for any company. Also, calculated loyalty scores enable segmentation at the customer level, providing a powerful framework for marketing teams, in order to address specific actions for each customer groups (or even individuals) and capture actionable insights about the business.

While our proposed method effectively models customer loyalty by combining established analytical techniques in a novel way, we recognize that the current evaluation is limited to a single dataset and business context. Future studies could aim to assess the robustness and scalability of the approach across multiple datasets from diverse industries, such as retail, finance and telecommunications. This would allow a more comprehensive understanding of how sector-specific dynamics may influence the clustering and scoring processes.

An additional opportunity would be expanding the comparative analysis to include other state-of-the-art loyalty prediction techniques, particularly those grounded in modern machine learning methods. This kind of comparison would provide a clearer benchmark about the performance of our model. Finally, future iterations of the model could also include integrating additional behavioral dimensions beyond RFM, such as channel usage, customer satisfaction indicators, and product category affinity, what should enhance its descriptive and predictive capabilities and supporting business decisions.

ACKNOWLEDGMENT

REFERENCES

[1] V. Shankar, "Big Data and Analytics in Retail," NIM Marketing Intelligence Review, vol. 11, number 1, pp. 36–40, 2019.

[2] Y. Chen, Y. C. Leong, L. S. Yiing, and Y. Xiao, "Study on the Influence of Knowledge-driven Technology on predicting consumer Repurchase Behaviour," International Journal of Communication Networks and Information Security, vol. 15, number 1, pp. 109–117, Eds. Kohat University of Science and Technology (KUST), 2023.

[3] A. Ehsani, and A. Hatamlou, "Identifying loyal customers and predicting customers purchase behavior using k-means and SOM algorithms," International Journal of Computing and Artificial Intelligence, vol. 1, number 1, pp. 21–27, 2020.

[4] A. G. Aggarwal, and S. Yadav, "Customer Segmentation Using Fuzzy-AHP and RFM Model," 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), pp. 77–80, 2020.

[5] D. Chen, K. Guo, and G. Ubakanma, "Predicting customer profitability over time based on RFM time series," International Journal of Business Forecasting and Marketing Intelligence, vol. 2, number 1, pp. 1–18, Eds. Inderscience Publishers (IEL), 2015.

[6] S. Chen, "Estimating customer lifetime value using machine learning techniques, chap. 2," Data Mining. IntechOpen, Rijeka, 2018.

[7] A. Matz, and A. T. Hermawan, "Customer Loyalty Clustering Model Using K-Means Algorithm with LRIFMQ Parameters," Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi, vol. 5, number 2, pp. 54–61, 2020.

[8] M. Khajvand, K. Zolfaghar, S. Ashoori, and S. Alizadeh, "Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study," Procedia computer science, vol. 3, pp. 57–63, Eds. Elsevier, 2011.

[9] F. Safari, N. Safari, and G. A. Montazer, "Customer lifetime value determination based on RFM model," Marketing Intelligence & Planning, vol. 34, number 4, pp. 446–461, Eds. Emerald Group Publishing Limited, 2016.

[10] S. Supangat, and Y. Mulyani, "Customer Loyalty Analysis Using Recency, Frequency, Monetary (RFM) and K-means Cluster for Labuan Bajo Souvenirs in Online Store," Journal of Information Systems and Informatics, vol. 5, number 1, pp. 285–299, 2023.

[11] R. Cao, X. Tang, C. Shi, Y. Wu, D. Tan, W. Liu, and M. Huang, "A RS-BOLGBM Algorithm Based on LightGBM for Supercomputing Center Job Running Status Prediction," Journal of Grid Computing, 2024.

[12] D. Chen, "Online Retail II [Dataset]". UCI Machine Learning Repository. https://doi.org/10.24432/C5CG6D. 2012.

[13] P. Anitha, and M. M. Patil, "RFM model for customer purchase behavior using K-Means algorithm," Journal of King Saud University – Computer and Information Sciences, vol. 34, pp. 1785–1792, 2022.

[14] J. Wu, L. Shi, L. Yang, X. Niu, Y. Li, X. Cui, S.-B. Tsai, and Y. Zhang, "User Value Identification Based on Improved RFM Model and K-Means++ Algorithm for Complex Data Analysis," Wireless Communications and Mobile Computing, pp. 1–8, 2021.

[15] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, "RFM ranking – an effective approach to customer segmentation," Journal of King Saud University - Computer and Information Sciences, vol. 33, pp. 1251–1257, 2021.

[16] M. A. Rahim, M. Mushafiq, S. Khan, and Z. A. Arain, "RFM-based repurchase behavior for customer classification and segmentation," Journal of Retailing and Consumer Services, vol. 61, 2021.