Integration of Sentiment Analysis and RFM in Restaurant Customer Segmentation: A 7P-Based CRM Model with Clustering

Budi Sunarko[1,*], Uswatun Hasanah[2], Syahroni Hidayat[3], Rina Rachmawati[4]

[1,2,3,4]Faculty of Engineering, Universitas Negeri Semarang, Sekaran, Semarang City 50229, Indonesia

(Received: January 5, 2025; Revised: February 7, 2025; Accepted: March 28, 2025; Available online: April 30, 2025)

Abstract

The increasing use of digital platforms like Tripadvisor has created opportunities to transform customer review data into strategic insights for Customer Relationship Management (CRM). This study proposes a novel CRM model by integrating the Recency, Frequency, Monetary (RFM) framework with the 7P marketing mix to segment restaurant customers more effectively. Using 3,716 Tripadvisor reviews, annotated based on 7P elements and clustered through unsupervised learning, three key customer segments were identified: acquisition, retention, and win-back. Evaluation metrics show strong clustering performance with a Silhouette Score of 0.73 and a Davies-Bouldin Score of 0.08. The acquisition cluster (Product) demonstrates the highest Frequency (37,664) and Monetary value (64.94), signifying high engagement and revenue potential. The retention cluster (Physical Evidence, Place, Process, Promotion, Traveler) shows stable interaction patterns with Recency values of 1261–1262 and moderate Frequency (378–2,079). The win-back cluster (Price, People) reflects lower Frequency (198–946) but equal Recency (1259), indicating recent but infrequent activity, which is ideal for reactivation strategies. By mapping customer reviews to 7P labels and analyzing them using RFM, the model uncovers specific behavioral patterns tied to service quality, pricing, and promotions. This integration allows restaurants to apply tailored strategies: offering loyalty rewards to high-frequency customers, promotional incentives for those with high Recency, and prioritizing high-monetary customers for exclusive programs. The novelty of this research lies in its combined use of sentiment-based review analysis and RFM–7P segmentation, offering a scalable, data-driven framework for enhancing customer satisfaction, loyalty, and long-term business growth in the restaurant industry.

Keywords: Customer Relationship Management, Sentiment Analysis, RFM (Recency, Frequency, Monetary), Clustering, User Reviews

1. Introduction

Information technology and big data development have driven significant transformations across various business sectors, particularly in implementing Customer Relationship Management (CRM) [1]. Initially, CRM functioned merely as a tool for managing customer data. Still, it has evolved into a technology-driven strategic approach designed to enhance customer retention, strengthen loyalty, and drive business growth [2]. One of the latest innovations in CRM is using data from digital platforms, such as Google Maps Reviews, to analyze customer sentiment and identify interaction patterns [3], [4]. Google Maps Reviews serve as a rich source of information, providing direct representations of customer experiences with a service or Product [5]. However, the unstructured nature of this data often poses challenges in optimizing its benefits without the appropriate analytical technologies.

Previous research has highlighted the importance of sentiment analysis in reinforcing CRM. Joung and Kim found that online review analysis can provide significant insights into product development and the personalization of customer services [6], [7]. Additionally, other studies emphasize the substantial potential of technology in supporting business strategies, particularly in the hospitality and restaurant sectors [8]. Sunarko et al. underscored the application of multi-class sentiment analysis using the BERT (Bidirectional Encoder Representations from Transformers) model to classify restaurant customer reviews into various categories such as menu, taste, and atmosphere [9]. Although the results demonstrated suboptimal accuracy (41.11%), this study affirms that integrating deep learning technology can enrich the understanding of customer experiences and lay the foundation for better strategy development in the restaurant industry. The 7P approach (Product, Price, Place, Promotion, People, Process, and Physical Evidence) in the marketing

***Corresponding author: Budi Sunarko (budi.sunarko@mail.unnes.ac.id)**
DOI: https://doi.org/10.47738/jads.v6i2.633

mix is crucial in understanding customer experiences. A study by Kwok et al. [10], on Airbnb customer reviews through big data and machine learning revealed that Service Product and Physical Evidence are the most frequently discussed elements, while Price and Promotion are less frequently mentioned [10]. This indicates that customers are more attentive to direct experiences and physical evidence when evaluating services, while price and promotional factors are often overlooked. Research by Zerbino et al. further supports this view, showing that integrating big data with CRM can enhance decision-making quality and strengthen customer relationships [11]. However, most previous studies have focused on text-based review analysis without profoundly considering the temporal aspects and Frequency of customer interactions.

Despite various studies integrating sentiment analysis and the 7P model into CRM, time-series sentiment analysis approaches combined with RFM and clustering remain rare [12]. This research aims to fill this gap by developing a more comprehensive CRM model for customer segmentation, providing more accurate solutions for understanding customer interaction patterns over time. To address the gap, this study integrates multi-aspect sentiment analysis and the 7P approach to understand customer behavior comprehensively.

Although various studies have incorporated sentiment analysis and the 7P model into CRM, the combination of time-series sentiment analysis with Recency, Frequency, Monetary (RFM), and clustering remains underexplored. This research seeks to overcome these limitations by developing a more comprehensive CRM model that leverages customer review data for segmentation, focusing not only on customer opinions but also on the temporal dimension and the intensity of their interactions. As a result, the model produced can enhance customer retention and identify strategic opportunities to attract new customers and recover those who have not interacted for a long time.

2. Related Work

With the advancement of digital technology and the growth of big data, many businesses have shifted their focus to leveraging sentiment analysis to enhance CRM. One of the latest trends is using customer reviews from online platforms such as Google Maps, TripAdvisor, and Airbnb to gain deeper insights into customer experiences. Various studies, such as those conducted by Sunarko et al. [9] and Kwok et al. [10], have demonstrated how customer review analysis can be applied to CRM based on the 7P model. However, this approach remains descriptive and has yet to integrate the temporal dimension and RFM-based segmentation, which is the primary focus of this research. Kwok et al. [10] and Sunarko et al. [9] applied sentiment analysis to classify keywords associated with each P. The K-Nearest Neighbors (KNN) and Ensemble classification methods were implemented to determine the confidence level of the relationship between keywords and each P based on their accuracy values. Consequently, this confidence level will enhance the reliability of the RFM analysis results.

The study by Sunarko et al. shows that applying multi-class text classification using the BERT model can classify restaurant customer reviews into various aspects such as menu, taste, and restaurant atmosphere. BERT was chosen because it is better at understanding context, which is needed in sentiment analysis. Although the accuracy achieved is not optimal, integrating deep learning technology into customer opinion analysis has opened new opportunities to understand customer needs more comprehensively [9].

Furthermore, the research by Kwok et al. highlights the application of the 7P marketing mix model as a practical framework for analyzing customer reviews on the Airbnb platform. The results indicate that Service Product and Physical Evidence are the primary focus in customer reviews, while Price and Promotion aspects are less emphasized. This suggests that customers place greater importance on direct experience and physical evidence as key indicators of their satisfaction with the service [10].

In the context of market segmentation, the study by Ahani et al. developed a hotel spa customer segmentation method through the analysis of reviews on TripAdvisor, utilizing a hybrid machine learning approach to predict customer travel choices [13]. This approach demonstrates how social data from customer reviews can be harnessed as business intelligence to support strategic decision-making. Additionally, the study by Fan et al. emphasizes how big data analytics has reshaped business intelligence through the lens of the marketing mix [14]. Companies can gain broader and more accurate insights to develop customer-oriented marketing strategies by analyzing data from various sources.

These studies demonstrate that integrating sentiment analysis, the 7P marketing mix approach, and big data-driven segmentation methods contribute significantly to developing more effective CRM models. By combining the RFM approach and clustering, this research aims to address the gap in the literature, which remains limited to time-based customer review analysis and direct interaction-based customer segmentation. The resulting CRM model is expected to strengthen retention, attract new customers, and recover inactive customers.

3. Methodology

This study adopts a data analysis approach based on sentiment analysis and customer segmentation using the RFM method integrated with clustering [12]. This process aims to develop a CRM model that identifies customer interaction patterns through review data. By employing machine learning techniques and sentiment analysis, this research focuses on segmenting customers based on the 7P aspects of the marketing mix [10], [15]. The study consists of the following key stages.

3.1. Data Collection and Preprocessing

The data collected for this study consists of customer reviews of restaurants. The dataset must include at least three attributes: User_ID, Review_Date, and Review_Content. The dataset used originates from Tripadvisor hotel reviews in New York, containing restaurant review data obtained from the data.world-website. This dataset comprises 3716 paragraphs representing restaurant reviews but lacks 7P labels, although it contains ratings. Therefore, it is necessary to annotate the reviews with 7P labels.

Upon obtaining the dataset, initial preprocessing steps were applied to refine the review data by removing symbols, extra spaces, and most punctuation, except for periods and commas. Removing unnecessary symbols, extra spaces, and punctuation in sentiment analysis helps reduce noise, improve tokenization, increase text consistency, optimize computational efficiency, and minimize bias in sentiment scores. This process also involved eliminating meaningless sentences resulting from incomplete initial sentences caused by imperfect data crawling. Only the review content was processed at this stage, while the User_ID and review date were left intact. The preprocessed data was then segmented to break down the original paragraph-shaped reviews into individual sentences, using periods (.) as delimiters. Each sentence retained its initial contextual information, including reviewer identity and review date.

3.2. Annotation and Sentiment Classification

During the annotation process, each sentence resulting from paragraph segmentation underwent labelling. A list of keywords was compiled for each P category to facilitate the annotation process, following a method similar to that used by Linchi Kwok [11]. The 7P keyword list used in this study was derived from Budi Sunarko's research on 7P classification, as shown in table 1. An additional Traveler label was included to reflect the user's perception, which was unrelated to the restaurant review. Further details can be referenced in the studies by Kwok and Budi.

The determination of annotation keywords for each of the 7Ps is based on the strength of their relationship. The Chi-square unigram method measures the strength of the association between keywords and the 7Ps. For each P, 12 keywords with the strongest relationships are generated. Subsequently, the results undergo a manual review process, where overly specific keywords, such as personal names, are eliminated. This ensures that only generalized keywords remain. The model's performance is expected to improve by using more generalized keywords.

Table 1. Keyword list for annotation under each 7P label

7P	Keywords
Product	tasty, highly, amazing, mozzarella, pasta, menu, pizza, fresh, delicious, food, drinks, beer
Price	fancy, value, reasonably, priced, worth, reasonable, prices, cheap, expensive, price
Place	old, close, place, hotel, near, walking, street, location
Promotion	come, return, again, back, would, recommended, will, definitely, highly, recommend
People	servers, his, owner, waitress, waiter, helpful, server, friendly, staff

Process	attentive, minutes, wait, fast, impeccable, greeted, reservation, table, service
Physical Evidence	seating, reviews, music, space, small, clean, beautiful, tables, cozy, atmosphere, outdoor
Traveler	we, my, visited, night, friends, birthday, family, went, wanted, visited

Next, sentiment scores were determined for each annotated review sentence. The annotation results and sentiment scores were used to assess each sentence's performance of the 7P labelling classification. This process mirrors the approach taken by Kwok and Budi S. The annotated data was subsequently transformed into a time-series dataset to enable RFM transformation and segmentation analysis. Figure 1 illustrates the steps to generate the time-series dataset.

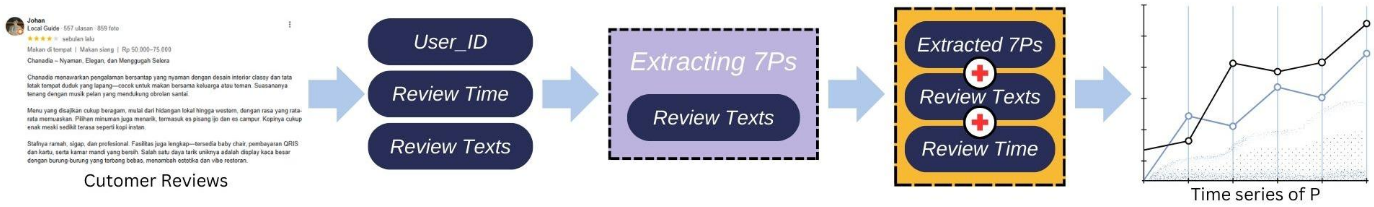

Figure 1. Constructing the Time Series Dataset

3.3. RFM Calculation and Data Normalization

The RFM approach is applied to transform the time-series sentiment dataset, which consists of categorical data, into numerical values. Recency measures how recently customers interacted with the restaurant through reviews, providing insights into their latest engagement. Frequency reflects the intensity of interaction, indicating how often a customer leaves reviews or returns for transactions. Meanwhile, Monetary evaluates the average contribution of customers based on the number of reviews, they submit during each interaction and the sentiment value of each review, representing the relative value of a customer to the business.

The implementation of RFM begins by grouping customer review data based on the 7P categories. Subsequently, data aggregation is performed for each group. During this process, Recency is calculated by subtracting the most recent review date from the current date. The result is then converted into the number of days, indicating the time elapsed since the customer's last interaction. A lower recency value signifies more recent customer interaction. Frequency is determined by summing all reviews within the group, representing the total number of interactions or reviews associated with a specific 7P element. A higher frequency value indicates that customers leave reviews more frequently. Monetary is derived by dividing the total number of reviews by the number of unique review dates. Monetary provides the average value of customer interaction, reflecting the extent of their contribution to each element.

The transformed RFM data is then normalized. The normalization process ensures that each variable in RFM has a uniform scale before clustering is performed. This step is crucial because each RFM component has different value ranges. For instance, Recency (number of days since the last interaction) may have a much larger scale than Frequency (number of reviews) or Monetary (average value per interaction). This imbalance can cause clustering algorithms such as K-Means to be skewed toward variables with the most enormous scale, resulting in suboptimal segmentation. The Z-score standardization method is applied for normalization. Each value is subtracted from the variable's mean and divided by its standard deviation. Z-score standardization is preferred for clustering tasks because it removes the influence of different feature scales, making distance-based algorithms like K-Means, DBSCAN, and Hierarchical Clustering more effective. Unlike Min-Max Scaling, it is more robust to outliers and maintains the distribution of data by centering it at a mean of 0 with a standard deviation of 1.

This normalization results in data with a mean of 0 and a standard deviation of 1. Normalization ensures that each RFM component has equal contributions during clustering, making the model more sensitive to actual patterns in the data without bias toward specific variables. Normalization enhances the effectiveness of clustering algorithms, producing more accurate and representative customer segmentation. It also improves model stability and convergence, especially for datasets with varying scales.

3.4. Clustering and Model Evaluation

At this stage, clustering and evaluation processes are conducted to group customers based on their normalized RFM values. Clustering is an unsupervised learning technique to identify patterns or groups within data without predefined labels. In this context, clustering is utilized for customer segmentation, allowing restaurants to group customers based on their behavioral characteristics and interaction patterns [16].

The first step in the clustering process is determining the optimal number of clusters. The code uses a cluster range starting from 2 up to the length of the normalized data. This approach ensures a minimum of two clusters, as clustering with a single cluster would not yield meaningful segmentation. An iterative process is conducted using the K-Means algorithm for each cluster count within the specified range [17]. K-Means cluster data into n clusters based on the nearest distance from the iteratively generated centroids. This process continues until the data points converge around specific centroids.

The clustering model evaluation uses two primary metrics: the Silhouette Score [18] and the Davies-Bouldin Score [19]. Silhouette Score and Davies-Bouldin Index are suitable for clustering evaluation without ground truth because they assess cluster quality based on compactness and separability. Silhouette Score evaluates how well each data point fits within its cluster. This score ranges from -1 to 1. A score closer to 1 indicates well-segmented data, while a score near -1 suggests poor clustering. Conversely, the Davies-Bouldin Score measures the distance between clusters and their compactness. A lower Davies-Bouldin Score is preferred, indicating that clusters are far apart and more compact within their groups. A comparison graph between the Silhouette and the Davies-Bouldin scores is generated to visualize model performance across different cluster counts. The number of clusters yielding the highest Silhouette Score is selected as the optimal cluster count. Subsequently, the K-Means model is applied using the chosen number of clusters.

After implementing the K-Means model, the clustering results are visualized using Principal Component Analysis (PCA) to reduce data dimensions into two principal components [20], [21]. This visualization allows high-dimensional data (three RFM variables) to be represented in two dimensions, facilitating interpretation. The visual map provides insights into cluster distribution and reveals distinct patterns among customer groups.

4. Results and Discussion

4.1. Research Findings

The research successfully clustered restaurant customer reviews based on the 7P values. This clustering resulted in three main clusters, each with different 7P elements, providing a foundation for implementing CRM recommendations for restaurant businesses, as shown in table 2.

Table 2. RFM Values by 7P and Cluster

7P	Cluster	Recency	Frequency	Monetary
People	2	1259	946	2.469974
Physical Evidence	0	1262	378	1.783019
Place	0	1261	470	1.850394
Price	2	1259	198	1.546875
Process	0	1261	737	2.219880
Product	1	1259	37664	64.93793
Promotion	0	1261	794	2.655518
Traveler	0	1261	2079	4.620000

Table 2 presents the RFM values categorized by 7P elements and cluster analysis results. Cluster 1 (Acquisition - Product) exhibits the highest Frequency of 37,664 and a significantly larger monetary value of 64.93 compared to other clusters. This underscores that the product factor is dominant in attracting new customers, with customers in this cluster engaging in highly frequent transactions and interactions. Cluster 0 (Retention - Physical Evidence, Place, Process, Promotion, Traveler) displays a more varied frequency distribution, ranging from 378 to 2,079, with a higher recency value (1261-1262). This pattern indicates that customers in this cluster tend to have stable yet less frequent interactions than those in the acquisition cluster. Cluster 2 (Win-Back - Price and People) shows significantly lower frequencies (946 and 198) but has the same recency value as the acquisition cluster (1259). This signifies that although the transaction frequency is low, customers in this cluster have recently left reviews, suggesting an opportunity to apply win-back strategies through pricing adjustments and service improvements. The table provides a clear depiction of each element's role in segmented CRM strategies, enabling a more targeted approach aligned with the characteristics of customers within each cluster.

Figure 2 illustrates the evaluation of clustering performance using the Silhouette Score and Davies-Bouldin Score to determine the optimal number of clusters. The analysis reveals that the highest Silhouette Score is achieved with three clusters, with a value of approximately 0.73, indicating well-separated clusters and optimal data distribution within each group. Meanwhile, the Davies-Bouldin Score achieves its lowest value (around 0.08) for clusters ranging from 4 to 6, suggesting that the clusters are sufficiently distant while maintaining high internal compactness. Based on these results, three clusters were chosen as the optimal configuration, striking the best balance between inter-cluster separation and intra-cluster compactness facilitating clearer and more defined customer segmentation.

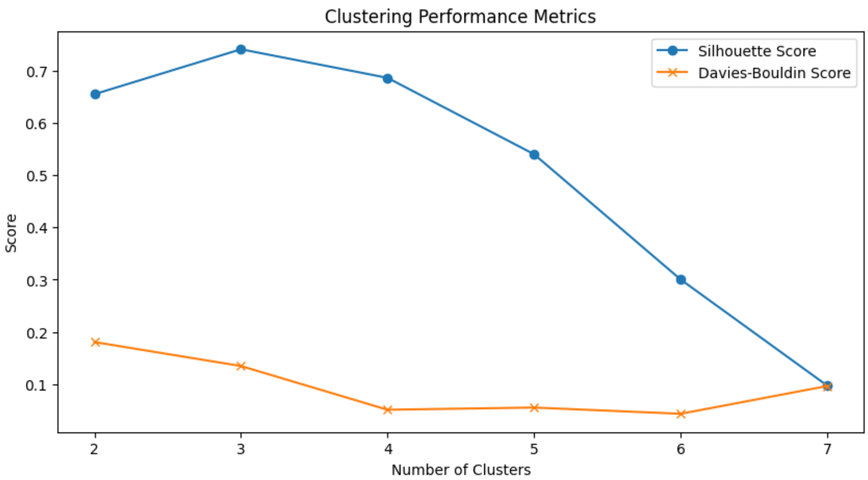

Figure 2. Clustering Performance by Silhouette and Davies-Bouldin Scores

Figure 3 visualizes the clustering results using PCA to reduce dimensionality and simplify the interpretation of data distribution in two-dimensional space. PCA is applied here to provide a more explicit representation of cluster patterns, allowing for easier identification of differences among customer groups.

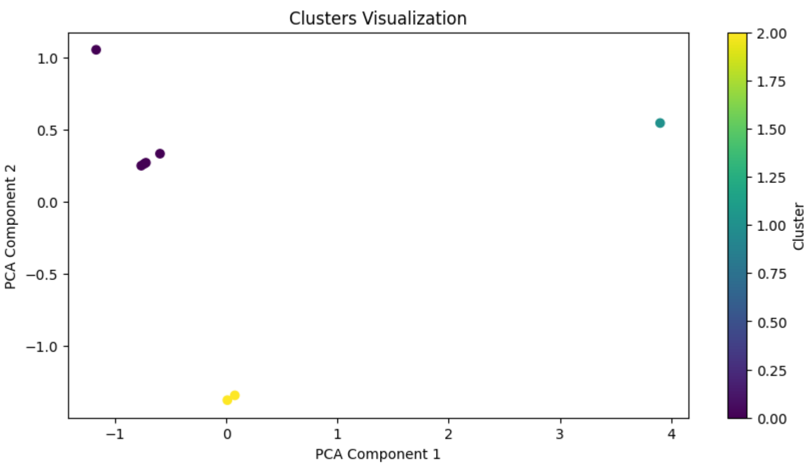

Figure 3. PCA Results for 7P Clusters

ISSN 2723-6471

In this study, PCA was utilized to enhance the visualization of clustering results by reducing the dimensionality of the dataset. Specifically, the PCA was applied to the scaled RFM dataset, which consists of customer segmentation features derived from customer reviews. The transformation process extracted two principal components, which serve as new representative features that retain the highest variance from the original RFM features while minimizing information loss. These two principal components were subsequently used to project the clustered data onto a two-dimensional space, enabling a clear graphical representation of cluster separability. The selection of two components for visualization was guided by the need to provide an interpretable and comprehensive depiction of the clustering structure, ensuring that the major patterns and relationships in the high-dimensional data were effectively captured. This visualization aids in assessing the quality of the clustering results, as well as in understanding the distribution and overlap of different customer segments.

The results reveal three primary clusters with distinct distributions. Cluster 1 (Acquisition Strategy - Product) stands out and is distinctly separated from the others, reflecting highly frequent interactions, indicating that the Product is central to attracting new customers. Cluster 0 (Retention Strategy), consisting of Physical Evidence, Place, Process, Promotion, and Traveler, appears more dispersed yet clustered, reflecting a focus on retaining existing customers. In contrast, Cluster 2 (Win-Back Strategy), covering Price and People, has a smaller and tighter distribution, suggesting a smaller cluster size but with significant potential to re-engage customers through pricing and service enhancements. This visualization aids in identifying appropriate strategic focuses for each customer segment based on their interaction patterns and sentiments.

Figure 4 illustrates the distribution of Frequency, Recency, and monetary values for each cluster, offering deep insights into customer interaction patterns. Cluster 1 (Acquisition - Product) stands out with the highest Frequency, reaching 37,664, indicating that customers in this group are highly active and contribute significantly to transactions and interactions. Additionally, this group has far higher Frequency and monetary values than other clusters, signifying that customer in this cluster are active contributors to revenue. In contrast, Cluster 0 (Retention - Physical Evidence, Place, Process, Promotion, Traveler) exhibits a more even distribution across various dimensions, reflecting stability in existing customer engagement. Meanwhile, Cluster 2 (Win-Back - Price and People) has significantly lower frequencies than other clusters but maintains the same Recency as Cluster 1 (1259). This suggests that although customers in this cluster rarely interact, they have left recent reviews, indicating an opportunity to re-engage them through targeted pricing and service improvements. This analysis reinforces the importance of segmentation in designing CRM strategies tailored to the characteristics and needs of each customer group.

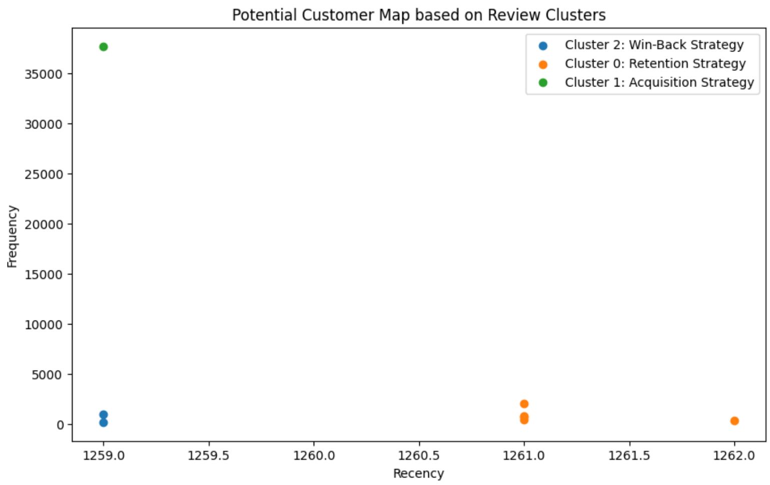

Figure 4. Customer Potential Map Based on Clustering Results

4.2. CRM Recommendations

Based on cluster analysis of restaurant customer review data, the implementation of CRM strategies integrated with the 7P model can be effectively carried out through win-back, retention, and acquisition strategies tailored to the characteristics of each cluster. This approach allows restaurants to design more personalized and relevant strategies to enhance customer satisfaction, retain loyalty, and expand market share.

 ISSN 2723-6471

Win-back strategies within Cluster 2 focus on Price and People, targeting efforts to recover customers who have disengaged or left negative reviews regarding pricing and service quality. Price, perceived as too expensive or uncompetitive, is often a key reason for customer attrition. Similarly, negative experiences with staff or subpar service exacerbate customer dissatisfaction. To address these issues, offering exclusive discounts or return promotions for customers who have not visited in a while can be beneficial. Staff training to enhance friendliness and service quality can also create more positive customer experiences. Loyalty programs that reward customers for referrals can further strengthen customer and restaurant staff relationships.

Retention strategies in Cluster 0 encompass Physical Evidence, Place, Process, Promotion, and Traveler. These strategies aim to maintain the satisfaction of active customers who have left positive reviews about the restaurant atmosphere, location, service efficiency, consistent promotions, and culinary experiences. Enhancing these elements can involve interior renovations, cleanliness, and improving seating comfort to elevate the overall customer experience. Strategically placed signage can enhance location visibility while optimizing reservation and online ordering systems can ensure seamless service. Continuous promotions, such as special events or seasonal discounts, can encourage repeat visits. Additionally, culinary tourism packages collaborating with hotels or travel agencies can attract travelers seeking unique dining experiences.

Acquisition strategies in Cluster 1 focus on Product. Innovation in products is critical for attracting new customers. Launching new menu items or enhancing popular products can drive acquisition efforts. Tasting events and social media promotions can increase visibility, drawing in potential customers. Collaborating with influencers or food bloggers can extend market reach and restaurant exposure. Additionally, offering special packages for first-time visitors, such as discounts for initial visits, can encourage new customer engagement.

By integrating win-back, retention, and acquisition strategies based on cluster analysis and the 7P model, restaurants can maximize customer satisfaction, improve loyalty, and expand their customer base. This approach ensures that each CRM element aligns with the specific needs of different customer segments, fostering more personalized and effective long-term relationships.

5. Conclusion

The study successfully developed a CRM model based on the 7P and RFM frameworks, effectively segmenting restaurant customers into three primary groups: acquisition, retention, and win-back. By leveraging customer review data, this model provides in-depth insights into customer interaction patterns, enabling restaurants to design more personalized and effective CRM strategies. Restaurant businesses are expected to adopt this approach to enhance customer loyalty and expand market share.

The RFM approach in customer data analysis allows restaurants to understand customer interaction patterns based on three key indicators comprehensively. Recency measures how recently customers have interacted with the restaurant through reviews, offering insights into their current engagement. Frequency reflects the intensity of interaction, indicating how often customers leave reviews or return for transactions. Meanwhile, Monetary assesses the average contribution of customers based on the number of reviews submitted per interaction, representing their relative value to the business.

In CRM, the 7P model helps identify distinct customer segments according to their interaction patterns and the value they provide. By categorizing review data by 7P labels and interaction dates, each element within the 7P framework can be analyzed to determine how actively customers respond to product quality, pricing, promotions, and other aspects. Applying RFM in CRM offers deep insights into which elements require greater attention, whether in retaining customers, attracting new ones, or recovering those at risk of churn.

Through this approach, restaurants can develop more personalized and effective strategies. Customers with low Recency and high Frequency can be directed towards loyalty programs to enhance their engagement. In contrast, those with high Recency and low Frequency can be targeted with special promotions to re-engage them. Additionally, customers with high monetary value can be prioritized for exclusive programs due to their significant contribution to

ISSN 2723-6471

the business. By applying this analysis, restaurants can manage customer relationships more strategically, ensuring that every customer interaction is factored into efforts to improve satisfaction and retention.

For future research, we will customize the study using real restaurants with more specific types, such as fine dining restaurants, fast casual restaurants, Asian cuisine, seafood restaurants, etc. Our goal is to obtain more detailed and measurable recommendations for each type of restaurant.

6. Declarations

6.1. Author Contributions

Conceptualization: B.S., U.H., S.H., and R.R.; Methodology: U.H. and S.H.; Software: B.S. and S.H.; Validation: B.S., U.H., and R.R.; Formal Analysis: B.S., U.H., S.H., and R.R.; Investigation: B.S.; Resources: S.H. and U.H.; Data Curation: U.H.; Writing Original Draft Preparation: B.S., U.H., S.H., and R.R.; Writing Review and Editing: U.H., B.S., and R.R.; Visualization: B.S. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

This research was supported by Directorate General of Higher Education, Research, and Technology (DRTPM) through the Fundamental Research scheme with contract number 070/E5/PG.02.00.PL/2024.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] G. J. Krishna and V. Ravi, "Evolutionary computing applied to customer relationship management: A survey," *Engineering Applications of Artificial Intelligence*, vol. 56, no.11, pp. 30–59, Nov. 2016, doi: 10.1016/j.engappai.2016.08.012.

[2] S. Lamrhari, H. E. Ghazi, M. Oubrich, and A. E. Faker, "A social CRM analytic framework for improving customer retention, acquisition, and conversion," *Technological Forecasting and Social Change*, vol. 174, no.1, pp. 1-19, Jan. 2022, doi: 10.1016/j.techfore.2021.121275.

[3] B. Mathayomchan and V. Taecharungroj, "'How was your meal?' Examining customer experience using Google maps reviews," *International Journal of Hospitality Management*, vol. 90, no.9, pp. 1-13, Sep. 2020, doi: 10.1016/j.ijhm.2020.102641.

[4] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P.-M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," *Expert Systems with Applications*, vol. 223, no.8, pp. 1-14, Aug. 2023, doi: 10.1016/j.eswa.2023.119862.

[5] B. Shin, S. Ryu, Y. Kim, and D. Kim, "Analysis on Review Data of Restaurants in Google Maps through Text Mining: Focusing on Sentiment Analysis," *J Multimed Inf Syst*, vol. 9, no. 1, pp. 61–68, Apr. 2022, doi: 10.33851/JMIS.2022.9.1.61.

[6] J. Joung and H. Kim, "Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews," *International Journal of Information Management*, vol. 70, no. 6, pp. 1-12, Jun. 2023, doi: 10.1016/j.ijinfomgt.2023.102641.

[7] K. I. Shaon, "A Theoretical Review of Crm Effects on Customer Satisfaction ond Loyalty," *Central European Business Review*, vol. 4, no. 01, pp. 23-36, 2015.

[8] C. Zhang, X. Wang, A. P. Cui, and S. Han, "Linking big data analytical intelligence to customer relationship management performance," *Industrial Marketing Management*, vol. 91, no.11, pp. 483–494, Nov. 2020, doi: 10.1016/j.indmarman.2020.10.012.

[9] B. Sunarko, U. Hasanah, and S. Hidayat, "Enhancing Restaurant Customer Review Analysis: Multi-Class Text Classification with BERT," in *2023 6th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Batam, Indonesia: IEEE, Dec. 2023, vol. 23, no. 12, pp. 501–506. doi: 10.1109/ISRITI60336.2023.10467438.

[10] L. Kwok, Y. Tang, and B. Yu, "The 7 Ps marketing mix of home-sharing services: Mining travelers' online reviews on Airbnb," *International Journal of Hospitality Management*, vol. 90, no. 9, pp. 1-11, Sep. 2020, doi: 10.1016/j.ijhm.2020.102616.

[11] P. Zerbino, D. Aloini, R. Dulmin, and V. Mininno, "Big Data-enabled Customer Relationship Management: A holistic approach," *Information Processing & Management*, vol. 54, no. 5, pp. 818–846, Sep. 2018, doi: 10.1016/j.ipm.2017.10.005.

[12] Ernawati, S. S. K. Baharin, F. Kasmin, and H. A. Purwanugraha, "Geo-marketing Promotional Target Selection using Modified RFM with Spatial and Temporal Analysis: A Case Study," J. Syst. Manag. Sci., vol. 12, no. 3, pp. 156–180, 2022.

[13] A. Ahani, M. Nilashi, O. Ibrahim, L. Sanzogni, and S. Weaven, "Market segmentation and travel choice prediction in Spa hotels through TripAdvisor's online reviews," *International Journal of Hospitality Management*, vol. 80, no. ?, pp. 52–77, Jul. 2019, doi: 10.1016/j.ijhm.2019.01.003.

[14] S. Fan, R. Y. K. Lau, and J. L. Zhao, "Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix," *Big Data Research*, vol. 2, no. 1, pp. 28–32, Mar. 2015, doi: 10.1016/j.bdr.2015.02.006.

[15] B. Sunarko, U. Hasanah, and S. Hidayat, "Comparison of Ensemble Learning Methods for Mining the Implementation of the 7 Ps Marketing Mix on TripAdvisor Restaurant Customer Review Data," *Int. J. Artif. Intell. Res*, vol. 7, no. 2, pp. 1-14, Dec. 2023, doi: 10.29099/ijair.v7i2.1096.

[16] A. Saxena, M. Prasad, A. Gupta, N. Bharill, O. P. Patel, A. Tiwari, M. J. Er, W. Ding, and C. T. Lin, "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, no. 12, pp. 664–681, Dec. 2017, doi: 10.1016/j.neucom.2017.06.053.

[17] L. E. Aik, T. W. Choon, and M. S. Abu, "K-means Algorithm Based on Flower Pollination Algorithm and Calinski-Harabasz Index," *J. Phys.: Conf. Ser.*, vol. 2643, no. 1, pp. 1-7, Nov. 2023, doi: 10.1088/1742-6596/2643/1/012019.

[18] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, no. 11, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7

[19] I. F. Ashari, R. Banjarnahor, D. R. Farida, S. P. Aisyah, A. P. Dewi, and N. Humaya, "Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies," *JAIC*, vol. 6, no. 1, pp. 07–15, Jul. 2022, doi: 10.30871/jaic.v6i1.3485.

[20] F. L. Gewers, G. R. Ferreira, H. F. D. Arruda, F. N. Silva, C. H. Comin, D. R. Amancio, and L. D. F. Costa, "Principal component analysis: A natural approach to data exploration," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–34, 2021.

[21] M. Greenacre, P. J. Groenen, T. Hastie, A. I. d'Enza, A. Markos, and E. Tuzhilina, "Principal component analysis," *Nature Reviews Methods Primers*, vol. 2, no. 1, pp. 1-100, 2022.