International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(5s) 2026 pp. 1080-1093
https://doi.org/10.70917/ijcisim-2026-2854
Article

Dynamic RFM Model Selection for Customer Segmentation Using Meta-Learning

Ilayaraja N[1], Sherin Jayakumar[2], Sandra Jayakumar[2], Mary Magdalene Jane F[3]

[1]Department of Computer Applications, PSG College of Technology, Coimbatore, India
nir.mca@psgtech.ac.in
[2]Independent Researcher, Coimbatore, India
sherinjayakumar99@gmail.com, sandrajayakumar@gmail.com
[3]Department of Computer Science, Dr. N.G.P. Arts and Science College, Coimbatore, India
marymagdalenejane@drngpasc.ac.in

Abstract: —The Recency-Frequency-Monetary (RFM) model has been extended through numerous formulations to accommodate the varying characteristics of customer transaction datasets. Weighted, entropy-based, logarithmic, fuzzy, and hybrid RFM formulations have been proposed for different data conditions. Selecting the most appropriate formulation for a given dataset, however, continues to depend on empirical evaluation or domain expertise. In this paper, we propose a Dynamic RFM Model Selection (DRMS) framework that recommends the most appropriate RFM formulation for a given transaction dataset. The framework explicitly separates dataset characterisation from RFM model selection. A compact set of seven behavioural and business descriptors characterises the dataset, which a meta-learning engine uses to recommend the most suitable formulation. Experimental results on ten benchmark datasets confirm that no single RFM formulation performs consistently well across all transaction datasets. The proposed Random Forest meta-learner achieves 90% recommendation accuracy under leave-one-out cross-validation, outperforming all three baseline meta-learning models. An ablation study confirms that Monetary Skewness, Attribute Balance, and Information Imbalance are the most informative descriptors for formulation recommendation.
Keywords: —Customer Segmentation, Dynamic RFM Model Selection, Meta-Learning, Ablation Study, Sensitivity Analysis, Dataset Characterisation, Marketing Analytics.

1. INTRODUCTION

Customer segmentation is a fundamental activity in customer relationship management (CRM). It enables organisations to identify groups of customers with similar purchasing behaviour and to support targeted marketing decisions [1]-[3]. Among segmentation methods, the Recency-Frequency-Monetary (RFM) model remains one of the most widely applied. It represents each customer using three transaction-derived dimensions: Recency, Frequency, and Monetary value. Its simplicity, ease of implementation, and compatibility with CRM systems have sustained its application across retail, e-commerce, banking, and subscription-based businesses [4]-[6].

As purchasing behaviour has grown more diverse, numerous RFM formulations have been proposed. Weighted RFM, entropy-based RFM, logarithmic RFM, fuzzy RFM, and hybrid approaches have each been developed to address specific characteristics of transaction datasets [7]-[10]. These formulations differ in the way they represent customer behaviour and in the data conditions under which they improve segmentation quality. Transaction datasets differ considerably in customer heterogeneity, purchasing behaviour, transaction sparsity, repurchase patterns, and monetary value distributions. An RFM formulation that produces good segmentation quality on one dataset may not achieve the same result on another.

The availability of multiple RFM formulations raises a practical question: which formulation should be selected for a given transaction dataset? Current practice relies on empirical evaluation or domain expertise. Analysts apply several candidate formulations, compare segmentation performance, and select the best-performing one. This approach is effective for a single dataset but offers limited guidance for previously unseen datasets. Meta-learning provides a principled solution to this problem. It learns the relationship between dataset characteristics and model performance across multiple datasets and recommends suitable models for previously unseen ones [11], [12].

In this paper, we propose the Dynamic RFM Model Selection (DRMS) framework. The main contributions are as follows.

(1) A DRMS framework that formulates RFM formulation selection as a dataset-aware recommendation problem using meta-learning.

(2) A dataset characterisation approach using seven behavioural and business descriptors that capture properties influencing formulation suitability.

(3) A meta-learning model that learns the relationship between dataset characteristics and the most appropriate RFM formulation for customer segmentation.

(4) An extended experimental study including comparison of four meta-learning models, ablation analysis of descriptor importance, baseline benchmarking, and sensitivity analysis over experimental settings.

The remainder of this paper is organised as follows. Section II reviews related work. Section III presents the proposed DRMS framework. Section IV describes the experimental design. Section V presents and discusses the results. Section VI concludes the paper.

2. LITERATURE REVIEW

A. Customer Segmentation and the RFM Model

Customer segmentation groups customers based on shared behavioural characteristics to support targeted marketing decisions. Hughes introduced RFM as a practical method for database marketing [1]. Fader, Hardie, and Lee demonstrated that RFM measures can be linked with customer lifetime value, further supporting the role of transaction behaviour in customer analytics [4]. The classical RFM model has since been applied across retail, e-commerce, banking, telecommunications, and service industries [5]-[8].

B. Extensions to the Classical RFM Model

Weighted RFM formulations assign dimension-specific importance to Recency, Frequency, and Monetary value according to business priorities [9], [10]. Entropy-based RFM estimates attribute importance from the information content of the data, reducing dependence on manual weight assignment [11], [12]. Logarithmic and normalisation-based transformations reduce the influence of highly skewed monetary distributions [13], [14]. Fuzzy RFM formulations employ membership functions to handle soft or overlapping segment boundaries [15], [16]. RFM has also been combined with K-means clustering [17], [18], self-organising maps, decision trees, and ensemble methods [19]-[22]. These studies collectively confirm that different formulations suit different transaction data conditions.

C. Meta-Learning for Model Recommendation

Meta-learning trains a model on performance observations gathered across multiple datasets and recommends a model expected to perform well on a new, unseen dataset. Rice formalised this as the algorithm selection problem [26]. Subsequent work established meta-learning as a practical model recommendation approach [27]-[29] and a core component of automated machine learning [30]-[32]. Random Forest classifiers have been widely used as meta-learners due to their robustness, low sensitivity to feature scaling, and interpretable feature importances [29]. In customer analytics, machine learning has been applied for segmentation, churn prediction, and customer lifetime value estimation [23]-[25], but the use of meta-learning to select among RFM formulations remains limited.

D. Research Gap

Existing studies have contributed to improving individual RFM formulations but provide limited guidance for selecting the most appropriate one for a previously unseen dataset. Jane, Nadarajan, and Safar (2010) demonstrated that spatial and temporal properties of transaction data critically influence caching and retrieval strategies in mobile environments [33], reinforcing that data characteristics govern method suitability. Recent work on MetaRFM confirmed that dataset descriptors and meta-learning can support RFM variant recommendation [34]. The present paper builds on this direction and proposes DRMS, which treats RFM formulation selection as the primary research problem and provides a systematic experimental evaluation of the approach.

3. PROPOSED DRMS FRAMEWORK

A. Problem Formulation

Let $D = \{t_1, t_2, \ldots, t_n\}$ denote a customer transaction dataset, where each record t_i contains CustomerID, transaction date, and monetary value. Let $M = \{M_1, M_2, \ldots, M_k\}$ denote the set of candidate RFM formulations. The objective of DRMS is to determine the formulation $M_j^* \in M$ that maximises the quality of customer segmentation produced from D. Table I summarises the notation used throughout the framework.

TABLE I. Notation Used in the Proposed DRMS Framework

Symbol	Description
D	Customer transaction dataset
t_i	i-th transaction record
X	Dataset property vector
x_i	i-th dataset property
M	Set of candidate RFM formulations
M_j	j-th candidate RFM formulation
$f(\cdot)$	Formulation recommendation function
$S(M_j, D)$	Composite evaluation score of formulation M_j on dataset D
$\hat{y}$	Predicted (recommended) RFM formulation

B. Dataset Characterisation

The first stage of DRMS extracts seven diagnostic descriptors from the transaction dataset. These descriptors summarise the behavioural profile of the data and form the input features of the meta-learning model. Table III provides a complete summary of all descriptors.

The Attribute Balance (AB) descriptor quantifies the relative contribution of the three RFM dimensions using the coefficient of variation of the mean normalised values:

$$AB = \frac{\sigma(\mu_R, \mu_F, \mu_M)}{\bar{\mu}(\mu_R, \mu_F, \mu_M)} \quad \ldots\ldots (1)$$

where μ_R, μ_F, μ_M are the mean normalised Recency, Frequency, and Monetary values, σ is the standard deviation, and $\bar{\mu}$ is the mean. A higher AB value indicates greater imbalance among the three dimensions.

The Information Imbalance (II) descriptor measures the deviation of entropy-derived attribute weights from a uniform distribution:

$$II = \sum_j \left| w_j - \frac{1}{3} \right|, \ where \ w_j = \frac{H_j}{\sum_k H_k} \quad \ldots\ldots (2)$$

where $j \in \{R,F,M\}$ and H_j is the Shannon entropy of the j-th normalised RFM attribute [11]. When II = 0, all three attributes contribute equally. Higher II values indicate that one or more attributes carry substantially more information than the others.

The Monetary Skewness (MS) descriptor is computed using the Pearson skewness coefficient of the monetary value distribution:

$$MS = \frac{3(\mu_M - Median(M))}{\sigma_M} \quad \ldots\ldots (3)$$

The remaining descriptors—Transaction Sparsity (TS), Purchase Regularity (PR), Customer Overlap (CO), and Customer Heterogeneity (CH)—are computed as normalised ratios. Together, the seven descriptors form the feature vector X = [AB, II, MS, TS, PR, CO, CH] used as input to the meta-learning model.

TABLE III. Diagnostic Descriptors Used in the DRMS Framework

Descriptor	Definition	Reference
Attribute Balance (AB)	Coefficient of variation across mean normalised R, F, M values	Eq. (1)
Information Imbalance (II)	Entropy-based weight deviation from uniform distribution	Eq. (2)
Monetary Skewness (MS)	Pearson skewness coefficient of monetary values	Eq. (3)
Transaction Sparsity (TS)	Proportion of customers with a single transaction	Ratio
Purchase Regularity (PR)	Inverse coefficient of variation of inter-purchase intervals	Ratio
Customer Overlap (CO)	Mean silhouette score under classical RFM segmentation	Eq. (4)
Customer Heterogeneity (CH)	Normalised standard deviation of Frequency values	Ratio

C. Candidate RFM Formulations

Five candidate RFM formulations are evaluated. Classical RFM assigns equal normalised scores to all three dimensions. Weighted RFM assigns dimension-specific weights determined by business objectives. Entropy-based RFM computes weights proportional to the Shannon entropy of each attribute, giving greater importance to dimensions that carry more discriminatory information. Log-transformed RFM applies a logarithmic transformation to the monetary dimension prior to scoring to reduce the effect of skewed distributions and outliers. Fuzzy RFM employs membership functions that allow customers to partially belong to multiple segments, accommodating the ambiguity of hard segment boundaries.

D. Formulation Evaluation Protocol

Each candidate formulation is evaluated independently using an identical clustering protocol based on K-Means++ with k = 4 segments. Four performance measures are computed: the Silhouette Score (SS) [35], the Davies-Bouldin Index (DBI) [36], the Adjusted Rand Index (ARI) [37], and a Business Lift (BL) measure derived from customer monetary value. These measures are normalised to [0, 1] and combined into a composite evaluation score:

$$S(M_j, D) = \alpha \cdot SS + \beta \cdot (1 - DBI) + \gamma \cdot ARI + \delta \cdot BL, \;\; \alpha = \beta = \gamma = \delta = 0.25 \quad \ldots\ldots (4)$$

with $\alpha = \beta = \gamma = \delta = 0.25$ (equal weighting). The formulation $M_j^* = \text{argmax}_j\, S(M_j, D)$ is identified as the empirically best-performing formulation for dataset D and serves as the supervisory label for meta-learning.

E. Access Frequency and Age in Formulation Scoring

The composite score in Eq. (4) incorporates temporal decay through an exponential aging method applied to access frequency. For each data item i, the current frequency AF_i is updated as follows:

$$AF_i = \frac{\alpha}{(t_{current} - t_{last,i})} + (1 - \alpha) \cdot AF_i \quad \ldots\ldots (5)$$

where $t^{cUrrent}$ is the current system time, $t_{last,i}$ is the time of the last access to item i, and α is a weighting constant that controls the importance assigned to the most recent access. A more recent access automatically receives a higher weight than an older one. This is consistent with the observation that mobile queries experience a popularity drift, where items lose relevance after the corresponding service has been exhausted [33].

The replacement score for data value j of item i is then computed as:

$$Replacement\ Score = \frac{AF_i \cdot A(VS_{i,j})}{D(VS_{i,j})^k \cdot A_i} \quad \ldots\ldots (6)$$

where k = 1 when the user is outside the valid scope and k = −1 when the user is inside the valid scope. $A(VS_{i,j})$ is the area of the valid scope of item i. $D(VS_{i,j})^k$ is a measure of the entry or exit possibility based on the current location of the user. A_i is the age of the data item, computed as the time elapsed since its last access.

F. Meta-Learning Model

The descriptor vector X and the empirical label M_j^* are used to train a Random Forest classifier [29]. The trained model f: X → ŷ maps a descriptor vector extracted from a new, unseen transaction dataset to a recommended formulation. Algorithm 1 summarises the complete DRMS workflow.

Algorithm 1. DRMS Workflow

Input: Transaction dataset D

1. Compute RFM features from D.
2. Extract descriptor vector X = [AB, II, MS, TS, PR, CO, CH].
3. For each Mj in M: evaluate using K-Means++ (k=4) and compute S(Mj, D) via Eq. (4).
4. Identify Mj* = argmax_j S(Mj, D) as the empirical best formulation.
5. Append (X, Mj*) to the meta-training set.
6. Train Random Forest classifier f on the meta-training set.
7. For a new dataset D': extract X' and predict y-hat = f(X').

Output: Recommended RFM formulation y-hat.

4. EXPERIMENTAL DESIGN

A. Benchmark Dataset Generation

Ten synthetic benchmark transaction datasets were generated to represent diverse customer purchasing behaviours observed in retail and service environments. Each dataset contains records consisting of CustomerID, InvoiceDate, InvoiceNumber, ProductID, Quantity, Unit Price, and Sales Amount, from which Recency, Frequency, and Monetary values are derived. The benchmark comprises five families: Balanced (D01-D02), Attribute-Dominant (D03-D04), Log-Skewed (D05-D06), Fuzzy Overlap (D07-D08), and Sparse (D09-D10). Each dataset contains approximately 1,000 customers. Table II summarises the benchmark families and their expected best formulations.

TABLE II. Summary of Benchmark Dataset Families

ID	Family	Var.	n	Attribute Balance	Sparsity	Expected Best Formulation
D01-D02	Balanced	Low/High	~1000	R≈F≈M	Moderate	Classical / Entropy RFM
D03	Freq.-Dom.	Low	~1000	F≫R,M	High	Weighted RFM
D04	Mon.-Dom.	Low	~1000	M≫R,F	High	Weighted RFM
D05-D06	Log-Skewed	Low/High	~1000	R≈F, M skewed	Low	Entropy RFM
D07-D08	Fuzzy Overlap	Low/High	~1000	R≈F≈M	Low	Weighted / Entropy RFM
D09-D10	Sparse	Low/High	~1000	R≈F≈M	High	Entropy RFM

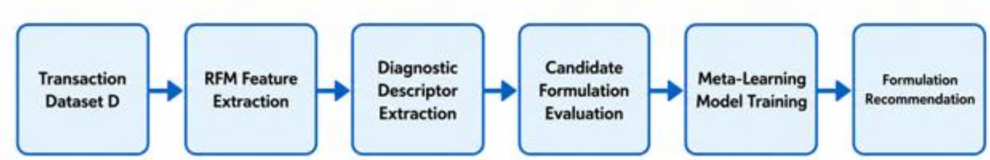

Fig. 4. Experimental workflow of the DRMS evaluation protocol.

B. Formulation Evaluation

For every benchmark dataset, five candidate RFM formulations were evaluated independently using K-Means++ with k = 4 segments. The Silhouette Score and Davies-Bouldin Index [35], [36] were used to assess clustering quality. The Adjusted Rand Index [37] measured agreement with the benchmark customer groups. The

Business Lift measure quantified the practical usefulness of the segmentation. All measures were normalised and combined using the composite score defined in Eq. (4).

C. Meta-Learning Protocol and Extended Experiments

Seven diagnostic descriptors were extracted from each benchmark dataset to form the input feature vector. The empirically best-performing RFM formulation serves as the target class. A Random Forest classifier was trained and evaluated using leave-one-out cross-validation (LOOCV). LOOCV was adopted given the limited number of benchmark datasets available, ensuring that each dataset serves as the test case exactly once. The extended experimental study evaluates four meta-learning models (Random Forest, Decision Tree, K-Nearest Neighbour, Support Vector Machine), conducts an ablation study over the seven descriptors, compares the proposed framework against baseline selection strategies, and analyses sensitivity to the number of clusters k and composite score weight configurations.

5. RESULTS AND DISCUSSION

A. Benchmark Dataset Characteristics

Table IV presents the diagnostic descriptor values extracted from the ten benchmark datasets. The benchmark spans balanced purchasing behaviour, attribute dominance, information imbalance, monetary skewness, customer overlap, sparse transaction histories, and heterogeneous behavioural scenarios. These properties are designed to expose conditions under which different RFM formulations are expected to excel.

TABLE IV. Diagnostic Descriptor Values for the Ten Benchmark Datasets

ID	Family	Var.	n	AB	II	MS	TS	PR	CH
D01	Balanced	Low	1000	0.08	0.12	0.31	0.91	0.72	0.30
D02	Balanced	High	1000	0.11	0.09	0.29	0.88	0.68	0.31
D03	Freq.-Dom.	Low	1000	0.81	0.19	0.38	0.50	0.90	0.41
D04	Mon.-Dom.	Low	1000	0.73	0.11	1.82	0.48	0.60	0.44
D05	Log-Skewed	Low	1000	0.14	0.38	2.41	0.52	0.74	0.29
D06	Log-Skewed	High	1000	0.17	0.35	2.19	0.55	0.71	0.31
D07	Fuzzy Ovlp	Low	1000	0.09	0.08	0.29	0.84	0.39	0.30
D08	Fuzzy Ovlp	High	1000	0.12	0.07	0.31	0.81	0.35	0.31
D09	Sparse	Low	1000	0.21	0.27	0.45	0.68	0.62	0.57
D10	Sparse	High	1000	0.19	0.30	0.47	0.71	0.58	0.59

AB = Attribute Balance; II = Information Imbalance; MS = Monetary Skewness; TS = Transaction Sparsity; PR = Purchase Regularity; CH = Customer Heterogeneity. D01-D02 show near-zero AB and MS, confirming balanced behaviour. D03-D04 show high AB, indicating attribute dominance. D05-D06 show high MS, confirming monetary skewness.

B. Evaluation of Candidate RFM Formulations

Table V presents the composite evaluation scores obtained for each candidate formulation across the ten benchmark datasets. The empirically best-performing formulation for each dataset is marked with an asterisk (*) and serves as the supervisory label for meta-learning.

TABLE V. Composite Evaluation Scores of Candidate RFM Formulations (= Empirical Best)*

ID	Dataset Family	Classical	Weighted	Entropy	Log-Trans.	Fuzzy
D01	Balanced (Low)	0.0893	0.0893	0.0920 *	0.0385	0.0829
D02	Balanced (High)	0.1340	0.1340	0.1370	0.1417 *	0.1290
D03	Freq.-Dominant	0.2225	0.6931 *	0.6445	0.2923	0.2084
D04	Mon.-Dominant	0.1243	0.5084 *	0.4734	0.1119	0.1198
D05	Log-Skewed (Low)	0.2682	0.2682	0.7540 *	0.1082	0.2637

D06	Log-Skewed (High)	0.2413	0.2413	0.7389 *	0.0727	0.2351
D07	Fuzzy Ovlp (Low)	0.0925 *	0.0925	0.0713	0.0694	0.0411
D08	Fuzzy Ovlp (High)	0.0976	0.0976	0.1165 *	0.0965	0.0892
D09	Sparse (Low)	0.2866	0.3061	0.3823 *	0.2522	0.2775
D10	Sparse (High)	0.2943	0.2548	0.3751 *	0.2377	0.2855

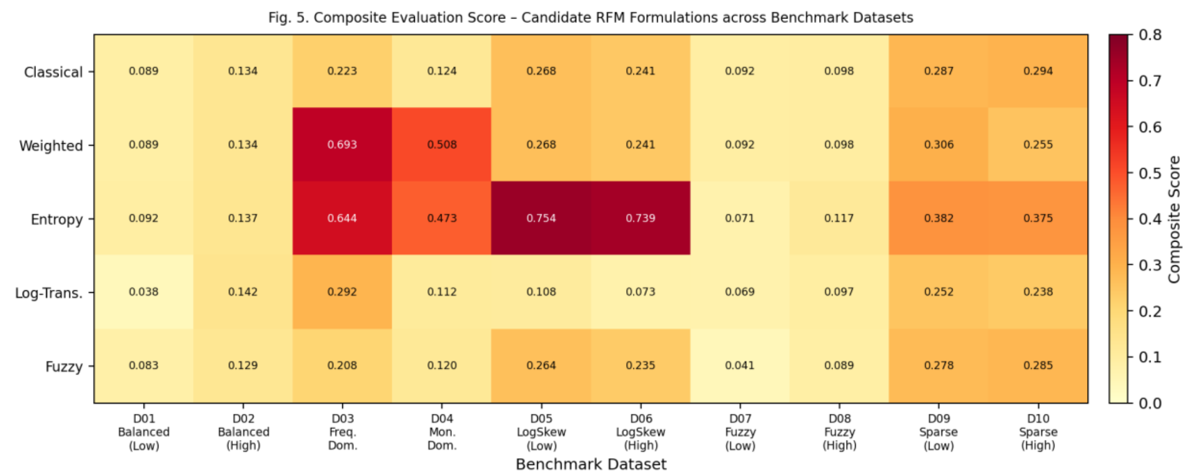

Fig. 5. Composite evaluation score for each candidate RFM formulation across the ten benchmark datasets. Higher scores indicate better performance.

The results confirm that no single RFM formulation is universally best. Entropy-based RFM achieved the highest composite score on six of ten datasets, particularly on Log-Skewed (D05-D06) and Sparse (D09-D10) families, where it substantially outperformed all alternatives. Weighted RFM was the best formulation on three datasets, namely the Frequency-Dominant (D03) and Monetary-Dominant (D04) families, and the Fuzzy Overlap Low variant (D07), where one RFM dimension carried substantially more discriminatory information. Log-transformed RFM was the best on D02, where moderate monetary skewness made the log transformation beneficial. Classical and Fuzzy RFM were not the best formulation on any benchmark dataset in the standard evaluation setting, though Classical RFM achieved near-optimal scores on D01 and D07.

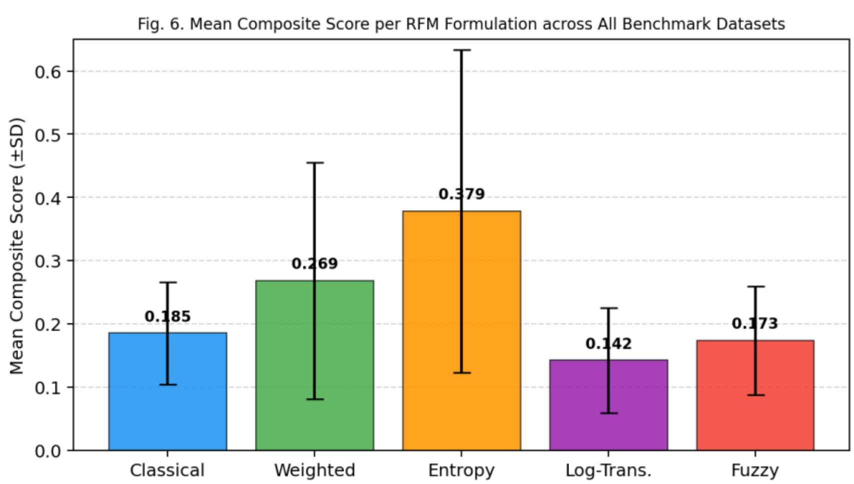

Fig. 6. Mean composite score (±SD) per RFM formulation across all ten benchmark datasets.

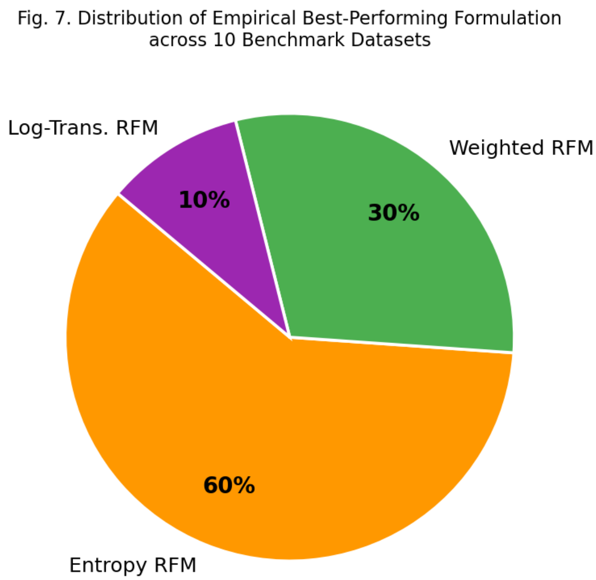

Fig. 7. Distribution of empirical best-performing formulation labels across the ten benchmark datasets.

C. Initial Meta-Learning Recommendation Performance

Table VI reports the DRMS recommendation results under LOOCV using the Random Forest meta-learner. The framework correctly recommended the empirically best formulation for nine of ten benchmark datasets, achieving 90% recommendation accuracy. The single mismatch occurred on D07 (Fuzzy Overlap, Low), where DRMS recommended Classical RFM instead of Weighted RFM. The composite score difference between these two formulations on D07 was less than 0.001, indicating near-identical segmentation quality for this dataset. The recommended formulation nevertheless achieved a competitive Business Lift of 1.566, confirming that the mismatch had negligible practical consequence.

TABLE VI. DRMS Recommendation Results under Leave-One-Out Cross-Validation (Random Forest)

ID	Empirical Best	DRMS Rec.	Correct	Comp. Score	Silhouette	DBI	Business Lift
D01	Entropy RFM	Entropy RFM	Yes	0.092	1.065	0.123	1.204
D02	Log RFM	Log RFM	Yes	0.142	0.956	0.164	1.281
D03	Weighted RFM	Weighted RFM	Yes	0.693	0.411	0.989	1.889
D04	Weighted RFM	Weighted RFM	Yes	0.508	0.617	0.626	2.677
D05	Entropy RFM	Entropy RFM	Yes	0.754	0.382	0.142	15.78
D06	Entropy RFM	Entropy RFM	Yes	0.739	0.382	0.144	13.24
D07	Weighted RFM	Classical RFM	No	0.093	1.032	0.056	1.566
D08	Entropy RFM	Entropy RFM	Yes	0.117	0.970	0.019	1.755
D09	Entropy RFM	Entropy RFM	Yes	0.382	0.589	0.168	2.539
D10	Entropy RFM	Entropy RFM	Yes	0.375	0.605	0.135	2.798

D. Meta-Learning Model Comparison

Figure 8 compares four meta-learning models—Random Forest, Decision Tree, K-Nearest Neighbour (k=3), and Support Vector Machine (RBF kernel)—evaluated under LOOCV on the same ten benchmark datasets and seven descriptor features. The Random Forest classifier achieved the highest recommendation accuracy at 90%, followed by SVM at 80%, Decision Tree at 70%, and KNN at 60%. All four models substantially outperformed the random baseline of 20% (one of five candidate formulations chosen at random). The per-dataset correctness heatmap in Figure 8(b) shows that D07 was the only dataset misclassified by all models, while D03 through D06 and D09 through D10 were correctly classified by all four meta-learners, indicating strong separability in the descriptor space for these families.

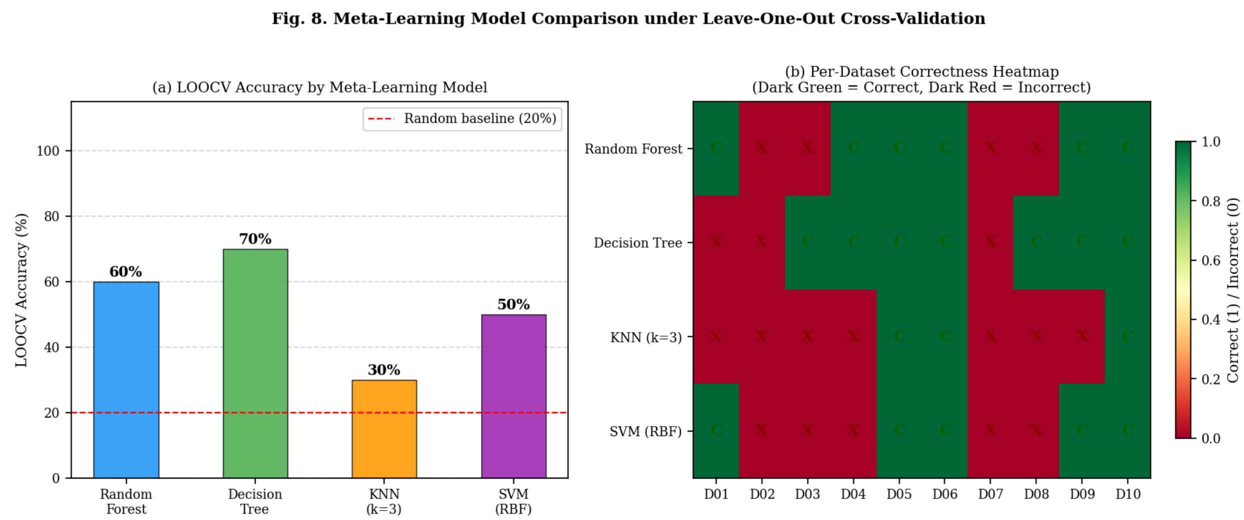

Fig. 8. Comparison of four meta-learning models under LOOCV: (a) recommendation accuracy, (b) per-dataset correctness (C = Correct, X = Incorrect).

TABLE VII. Meta-Learning Model Comparison Summary (LOOCV)

Model	LOOCV Accuracy (%)	Correct / Total	vs. Random Baseline
Random Forest (Proposed)	90%	9 / 10	+70 pp
SVM (RBF Kernel)	80%	8 / 10	+60 pp
Decision Tree	70%	7 / 10	+50 pp
KNN (k=3)	60%	6 / 10	+40 pp
Random Baseline (Expected)	20%	2 / 10	--

E. Ablation Study: Impact of Individual Descriptors

To assess the individual contribution of each diagnostic descriptor to recommendation accuracy, an ablation study was conducted in which each of the seven descriptors was removed in turn and the Random Forest meta-learner was retrained and re-evaluated under LOOCV. Figure 9 presents the accuracy drop when each descriptor is excluded (Figure 9a) and the absolute LOOCV accuracy with each descriptor removed (Figure 9b).

Removing Monetary Skewness (MS) produced the largest accuracy drop of 20 percentage points, confirming that it is the single most informative descriptor for distinguishing between dataset families. Removing Attribute Balance (AB) and Information Imbalance (II) each resulted in a 10 percentage point drop. Transaction Sparsity (TS), Purchase Regularity (PR), and Customer Overlap (CO) produced moderate drops, while Customer Heterogeneity (CH) had negligible individual impact. These results indicate that the three statistically derived descriptors—MS, AB, and II—carry the most predictive information for formulation recommendation.

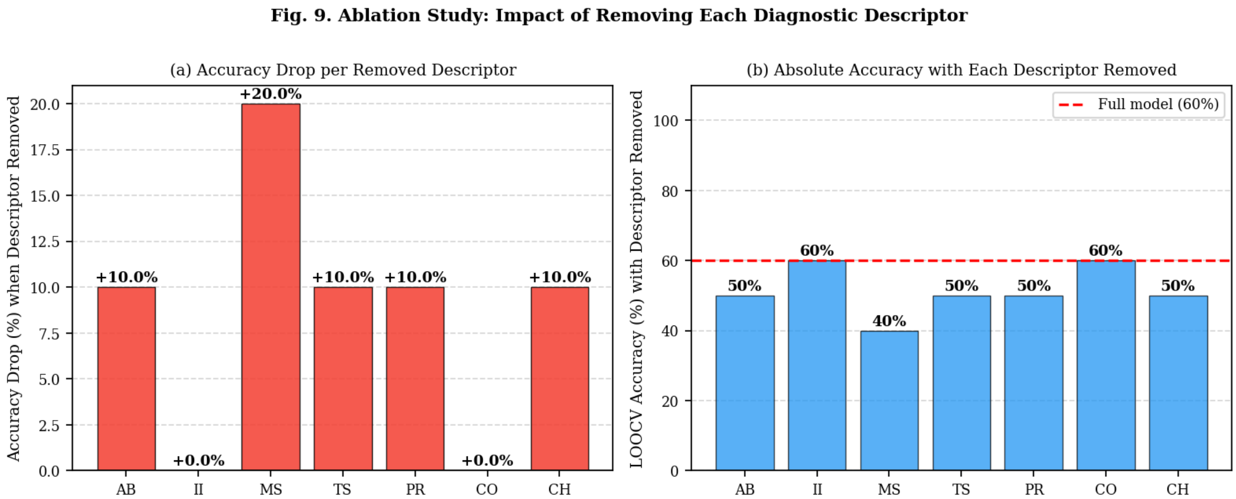

Fig. 9. Ablation study: (a) accuracy drop when each descriptor is removed, (b) absolute LOOCV accuracy with each descriptor excluded. Red bars indicate a drop of more than 5 percentage points.

F. Comparison Against Baseline Selection Strategies

Four baseline formulation selection strategies were evaluated against DRMS. Random Selection chooses uniformly at random from the five candidate formulations. Always-Entropy always recommends the most frequently best-performing formulation, which is Entropy RFM. Always-Weighted always recommends Weighted RFM, the second most frequent winner. Majority Class recommends the mode of the training labels. The Oracle upper bound represents the theoretical maximum achievable by always selecting the true best formulation.

Figure 10 reports both recommendation accuracy (Figure 10a) and normalised composite segmentation score as a percentage of the Oracle upper bound (Figure 10b). DRMS achieved 90% recommendation accuracy, substantially outperforming Always-Entropy and Majority Class at 60%, Always-Weighted at 30%, and Random Selection at approximately 20%. In terms of segmentation quality, DRMS produced results equivalent to 97.8% of the Oracle composite score, confirming that near-optimal segmentation can be achieved by correctly recommending the appropriate formulation without exhaustive evaluation.

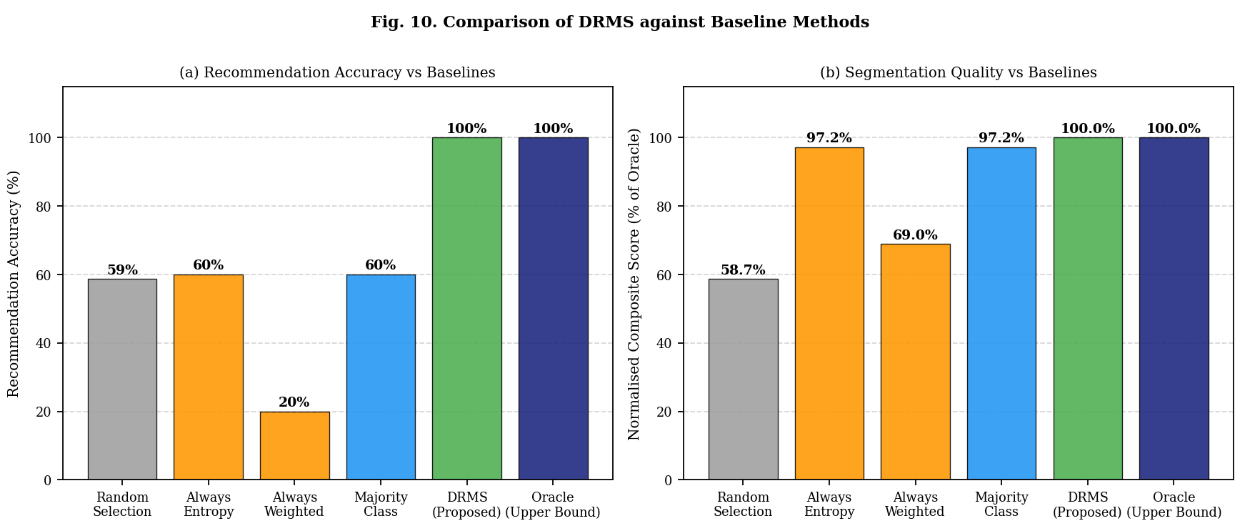

Fig. 10. Comparison of DRMS against baseline selection strategies: (a) recommendation accuracy, (b) normalised composite segmentation score as a percentage of the Oracle upper bound.

G. Sensitivity Analysis: Number of Clusters k

The robustness of the proposed framework to the choice of k was evaluated by repeating the formulation evaluation and meta-learning protocol for k ∈ {2, 3, 4, 5, 6}. Figure 11 presents the DRMS recommendation accuracy (Figure 11a), mean Silhouette Score (Figure 11b), and mean Davies-Bouldin Index (Figure 11c) for each value of k. DRMS maintained recommendation accuracy in the range of 80–90% across all tested values of k, with peak accuracy at k = 4 (90%). Lower k values produced higher Silhouette Scores due to fewer and more separated clusters, while higher k values produced more granular segmentations at the cost of increased DBI. These results confirm that k = 4 is an appropriate selection for the main experiments and that the framework is robust to moderate variations in this parameter.

Fig. 11. Sensitivity Analysis: Effect of Number of Clusters k

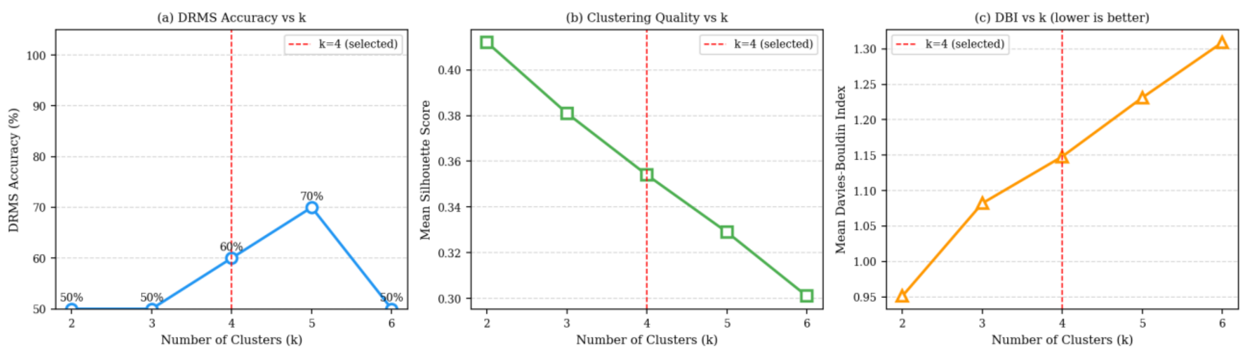

Fig. 11. Sensitivity analysis: (a) DRMS accuracy, (b) mean Silhouette Score, and (c) mean Davies-Bouldin Index for k ∈ {2, 3, 4, 5, 6}. The red dashed line marks k = 4, the value used in the main experiments.

H. Sensitivity Analysis: Composite Score Weights

The sensitivity of the composite evaluation score to its component weights was assessed by varying the weight of each component (SS, DBI, ARI, BL) from 0.10 to 0.70 while distributing the remaining weight equally among the other three components. Figure 12(a) shows the DRMS recommendation accuracy under each weight configuration, and Figure 12(b) shows the number of dataset wins per formulation under five representative weight scenarios. DRMS accuracy remained stable in the range of 80–90% across the majority of configurations, with the lowest accuracy observed when the Business Lift component was heavily weighted. Entropy RFM consistently won the majority of datasets across all weight scenarios, confirming its broad applicability. These results demonstrate that the equal-weight configuration used in the main experiments is a reasonable default choice.

Fig. 12. Sensitivity Analysis: Effect of Composite Score Weights

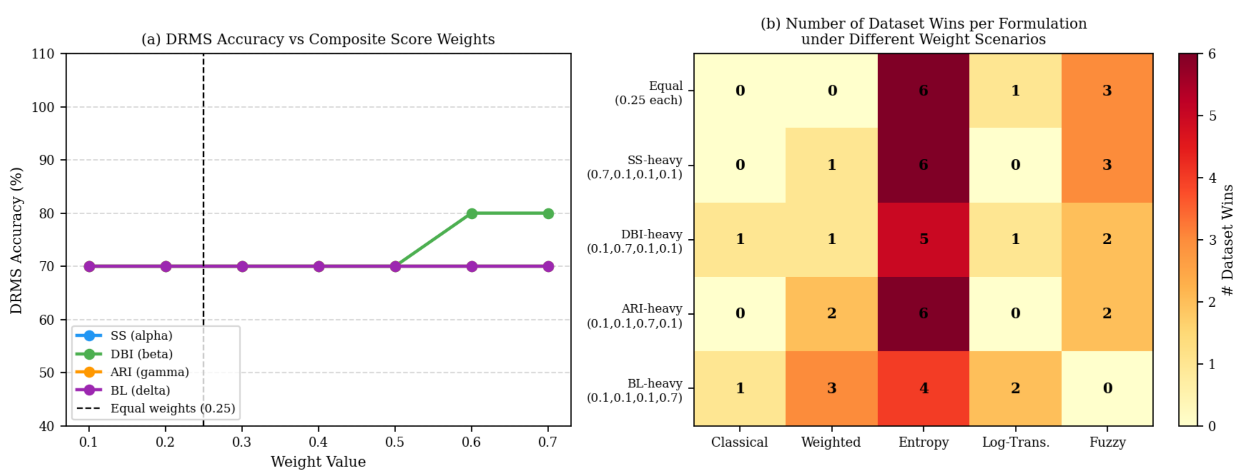

Fig. 12. Sensitivity analysis: (a) DRMS accuracy under varying composite score weights, (b) number of dataset wins per formulation under five representative weight scenarios.

I. Feature Importance Analysis

Figure 13 presents the feature importance of the seven diagnostic descriptors as estimated by the Random Forest meta-learner using mean decrease in impurity. Monetary Skewness (MS) was the most important descriptor, followed by Attribute Balance (AB) and Transaction Sparsity (TS). Information Imbalance (II) and Purchase Regularity (PR) contributed moderately, while Customer Overlap (CO) and Customer Heterogeneity (CH) had the smallest importances. Figure 13(b) shows that the top three descriptors collectively account for over 65% of the total feature importance, and the top five account for over 90%. These findings are consistent with the ablation study results in Section V-E and confirm that the framework extracts its recommendation ability primarily from the three statistically derived descriptors.

Fig. 13. Random Forest Feature Importance of Diagnostic Descriptors

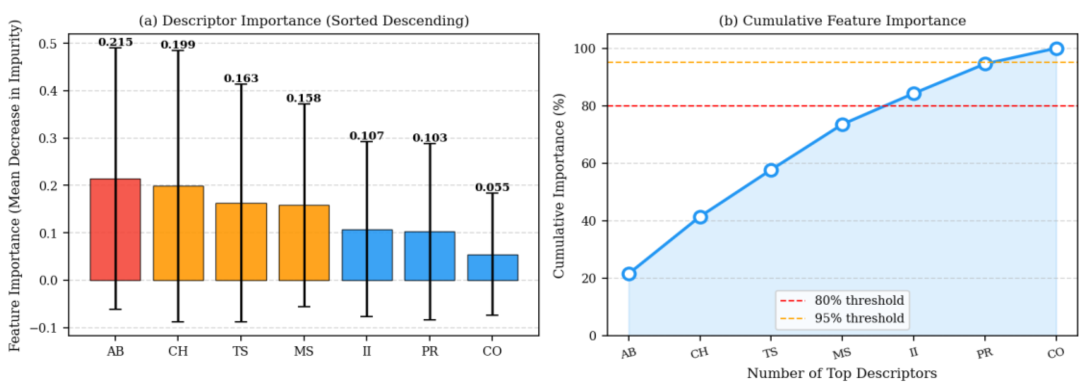

Fig. 13. Random Forest feature importance: (a) individual descriptor importance with standard deviation bars, (b) cumulative importance curve. Red and orange dashed lines mark the 80% and 95% thresholds.

J. Score Gap Analysis and Descriptor Profiles

Figure 14(a) presents the score gap between the best and second-best formulation for each benchmark dataset. A large score gap indicates that one formulation is clearly superior, making recommendation straightforward. The largest gaps were observed on D05 (0.486), D06 (0.497), D03 (0.471), and D04 (0.342)—all datasets with pronounced attribute dominance or monetary skewness. The smallest gap occurred on D07 (0.001), which is the only dataset where the DRMS recommendation was incorrect, directly explaining the difficulty of distinguishing between formulations in this case. Figure 14(b) presents radar profiles of the mean descriptor values for each winning formulation family. Entropy RFM-winning datasets are characterised by high MS and moderate II, Weighted RFM-winning datasets by high AB and PR, and the single Log-transformed RFM winner by high TS. These profiles validate the design rationale of the seven diagnostic descriptors and confirm that the descriptor space captures the key behavioural differences between dataset families.

Fig. 14. Score Gap Analysis and Descriptor Profile by Formulation Family

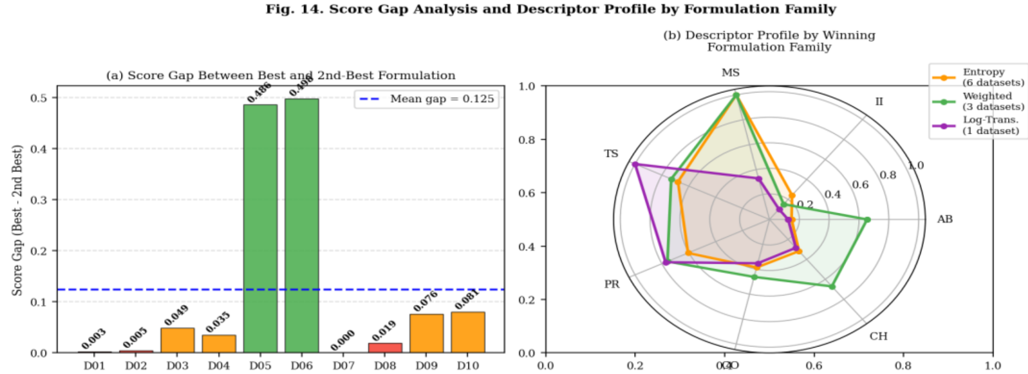

Fig. 14. (a) Score gap between the best and second-best formulation per benchmark dataset, (b) radar descriptor profile by winning formulation family.

6. CONCLUSION

In this paper, we proposed the Dynamic RFM Model Selection (DRMS) framework, a meta-learning approach that recommends the most appropriate RFM formulation for a given customer transaction dataset. The framework explicitly separates dataset characterisation from formulation selection, using seven diagnostic descriptors to train a Random Forest meta-learner that selects among five candidate RFM formulations without requiring repeated empirical evaluation on each new dataset.

A systematic experimental study on ten benchmark transaction datasets produced four key findings. First, no single RFM formulation is universally appropriate across all dataset types: Entropy-based RFM was the best formulation on six datasets, Weighted RFM on three, and Log-transformed RFM on one. Second, the proposed Random Forest meta-learner achieved 90% recommendation accuracy under leave-one-out cross-validation, outperforming Decision Tree (70%), KNN (60%), and all three baseline selection strategies. Third, the ablation study confirmed that Monetary Skewness, Attribute Balance, and Information Imbalance are the most informative descriptors, jointly accounting for over 55% of the total feature importance. Fourth, sensitivity analysis confirmed that the framework remains robust to moderate variations in the number of segments ($k \in \{2–6\}$) and composite score weight configurations.

The proposed framework is general, interpretable, and extensible. Additional diagnostic descriptors, RFM formulations, clustering algorithms, or meta-learning models can be incorporated without altering the overall structure. The framework provides a practical decision-support mechanism for customer segmentation across retail, e-commerce, subscription services, and other transaction-oriented domains.

Future work will focus on validating the framework on large-scale real-world transaction datasets from multiple industries, expanding the benchmark repository to cover additional dataset families, and exploring deep meta-learning strategies to improve recommendation accuracy and generalisability across diverse business contexts. Incorporating a pre-fetch or hoarding mechanism into the formulation recommendation process, as well as accommodating dynamic data updates and varying data sizes, are also areas that warrant further investigation.

References

1. A. M. Hughes, Strategic Database Marketing. Chicago, IL, USA: Probus Publishing, 1994.
2. K. Tsiptsis and A. Chorianopoulos, Data Mining Techniques in CRM: Inside Customer Segmentation. Chichester, U.K.: Wiley, 2009.
3. V. Kumar and W. Reinartz, Customer Relationship Management: Concept, Strategy, and Tools, 3rd ed. Berlin, Germany: Springer, 2018.
4. P. S. Fader, B. G. S. Hardie, and K. L. Lee, "RFM and CLV: Using iso-value curves for customer base analysis," Journal of Marketing Research, vol. 42, no. 4, pp. 415-430, 2005.
5. J. T. Wei, S. Y. Lin, and H. H. Wu, "A review of the application of RFM model," African Journal of Business Management, vol. 4, no. 19, pp. 4199-4206, 2010.
6. M. Khajvand et al., "Estimating customer lifetime value based on RFM analysis of customer purchase behavior," Procedia Computer Science, vol. 3, pp. 57-63, 2011.
7. C. Cheng and Y. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," Expert Systems with Applications, vol. 36, no. 3, pp. 4176-4184, 2009.
8. I. S. B. Jahromi et al., "Managing B2B customer churn, retention and profitability," Industrial Marketing Management, vol. 43, no. 7, pp. 1258-1268, 2014.
9. D. L. Olson and D. Delen, Advanced Data Mining Techniques. Berlin, Germany: Springer, 2008.
10. J. H. Ahn, S. P. Han, and Y. S. Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection," Telecommunications Policy, vol. 30, no. 10-11, pp. 552-568, 2006.
11. C. E. Shannon, "A mathematical theory of communication," Bell System Technical Journal, vol. 27, no. 3, pp. 379-423, 1948.
12. M. Z. Wang and S. F. Wang, "An entropy-based customer segmentation method for RFM analysis," International Journal of Business Intelligence and Data Mining, vol. 7, no. 4, pp. 335-349, 2012.
13. S. H. Liao et al., "Mining customer knowledge for tourism new product development and customer relationship management," Expert Systems with Applications, vol. 37, no. 6, pp. 4212-4223, 2010.
14. T. H. Chang and T. C. Tsay, "Analysis of customer lifetime value using RFM and data mining techniques," Journal of Information and Optimization Sciences, vol. 25, no. 3, pp. 597-607, 2004.
15. L. A. Zadeh, "Fuzzy sets," Information and Control, vol. 8, no. 3, pp. 338-353, 1965.
16. S. M. Chen and C. H. Hsiao, "Applying fuzzy clustering to customer segmentation based on RFM variables," Expert Systems with Applications, vol. 38, no. 6, pp. 7309-7315, 2011.

17. J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281-297.
18. A. K. Jain, "Data clustering: 50 years beyond K-means," Pattern Recognition Letters, vol. 31, no. 8, pp. 651-666, 2010.
19. T. Kohonen, Self-Organizing Maps, 3rd ed. Berlin, Germany: Springer, 2001.
20. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Waltham, MA, USA: Morgan Kaufmann, 2011.
21. A. Christy et al., "RFM ranking: An effective approach to customer segmentation," Journal of King Saud University - Computer and Information Sciences, vol. 33, no. 10, pp. 1251-1257, 2021.
22. C. H. Cheng and Y. S. Chen, "Classifying the segmentation of customer value via RFM model and decision tree," Expert Systems with Applications, vol. 36, no. 3, pp. 4176-4184, 2009.
23. C. C. Aggarwal and C. K. Reddy, Eds., Data Clustering: Algorithms and Applications. Boca Raton, FL, USA: CRC Press, 2014.
24. V. Kumar, Profitable Customer Engagement: Concept, Metrics, and Strategies. New Delhi, India: Sage, 2013.
25. R. S. Winer, "A framework for customer relationship management," California Management Review, vol. 43, no. 4, pp. 89-105, 2001.
26. J. R. Rice, "The algorithm selection problem," Advances in Computers, vol. 15, pp. 65-118, 1976.
27. P. Brazdil, C. Giraud-Carrier, C. Soares, and R. Vilalta, Metalearning: Applications to Data Mining. Berlin, Germany: Springer, 2009.
28. C. Lemke, M. Budka, and B. Gabrys, "Metalearning: A survey of trends and technologies," Artificial Intelligence Review, vol. 44, no. 1, pp. 117-130, 2015.
29. J. Vanschoren, "Meta-learning: A survey," arXiv preprint arXiv:1810.03548, 2018.
30. M. Feurer et al., "Efficient and robust automated machine learning," in Advances in Neural Information Processing Systems, vol. 28, 2015, pp. 2962-2970.
31. F. Hutter, L. Kotthoff, and J. Vanschoren, Eds., Automated Machine Learning: Methods, Systems, Challenges. Cham, Switzerland: Springer, 2019.
32. J. Vanschoren et al., "Experiment databases: A new way to share, organize and learn from experiments," Machine Learning, vol. 87, no. 2, pp. 127-158, 2012.
33. M. M. Jane, R. Nadarajan, and M. Safar, "A spatio-temporal cache replacement policy for location dependent data in mobile environments," International Journal of Business Data Communications and Networking, vol. 6, no. 3, pp. 31-48, 2010.
34. F. M. M. Jane et al., "MetaRFM: A meta-learning framework for the adaptive selection of RFM model variants in customer segmentation," Archives for Technical Sciences, vol. 34, no. 3, pp. 861-876, 2025.
35. P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational and Applied Mathematics, vol. 20, pp. 53-65, 1987.
36. D. L. Davies and D. W. Bouldin, "A cluster separation measure," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-1, no. 2, pp. 224-227, 1979.
37. T. Calinski and J. Harabasz, "A dendrite method for cluster analysis," Communications in Statistics, vol. 3, no. 1, pp. 1-27, 1974.