

# Integration of Decision-Making Method and Data-Mining Method as A Preliminary Study of Novel Sales Forecasting Method

Rendra Gustriansyah<sup>1,2</sup>, Ermatita<sup>3</sup>, Dian Palupi Rini<sup>4</sup>, Reza Firsandaya Malik<sup>5</sup>

<sup>1</sup> Faculty of Computer Science, Universitas Indo Global Mandiri, Indonesia, [rendra@uigm.ac.id](mailto:rendra@uigm.ac.id)

<sup>2</sup> PhD Student of Informatics Engineering, Faculty of Engineering, Universitas Sriwijaya, Indonesia

<sup>3</sup> Faculty of Computer Science, Universitas Sriwijaya, Indonesia, [ermatitaz@yahoo.com](mailto:ermatitaz@yahoo.com)

<sup>4</sup> Faculty of Computer Science, Universitas Sriwijaya, Indonesia, [dprini@unsri.ac.id](mailto:dprini@unsri.ac.id)

<sup>5</sup> Faculty of Computer Science, Universitas Sriwijaya, Indonesia, [rezafm@unsri.ac.id](mailto:rezafm@unsri.ac.id)



## ABSTRACT

Sales forecasting is intended to control the amount of available stock so that shortages or excess stock can be minimized, the fulfillment of consumer demand can be prepared on time, and cooperation with suppliers can be maintained properly. Therefore, this study aims to develop a new sales forecasting method that integrates the concept of RFM, the data-mining method, and the Best-Worst Method (BWM), where RFM is a strong-method in the field of database marketing and BWM is a new decision-making method. This research is a preliminary study that outlines the theory and phases of the proposed sales forecasting method. Based on the concept offered, research on the proposed sales forecasting method is very possible to be developed. This new sales forecasting method can be an alternative method for company management to optimize services to consumers, inventory efficiency, and predict the company's economic benefits in the future.

**Key words:** BWM, data-mining, decision-making, RFM, sales forecasting.

## 1. INTRODUCTION

In today's competitive market, sales forecasting plays an important role in business planning because it can minimize risk in decision making so the company can operate efficiently.

Sales forecasting is part of the management function. Companies that sell hundreds or even thousands of product items must be able to anticipate consumer demand. Because the company's failure to provide products will have a negative impact on the level of service to consumers. As a result, company profits will decrease because consumers will choose companies that can meet their needs.

For this reason, company management must be able to plan consumer demand starting from a few months before, because of the lead time for the procurement of products from suppliers.

Sales forecasting is not a simple refill but effective management and control to increase market penetration capacity, optimization of services to consumers, and efficiency for companies.

Inaccurate sales forecasting can increase costs associated with inventory so that investment becomes inefficient. To answer this problem, this research will propose a new sales forecasting method as a preliminary study [1], [2] that integrates the RFM concept, the data-mining clustering method (kMeans) [3], and the Best-Worst Method, where RFM is a powerful-method in marketing and BWM is a new decision-making method [4], [5].

Based on the literature obtained from Scopus in the past 5 years; there are 507 articles/proceeding papers/chapters related to sales forecasting, only 22 articles/chapters that discuss the use of RFM in sales forecasting, and no research is found that addresses the use of RFM and BWM in the area of sales forecasting. The combination of these two approaches is a new method in the field of sales forecasting which is expected to produce more accurate forecasting results, also useful as a literature review for future research.

## 2. LITERATURE REVIEW

Recency, frequency, and monetary (RFM) were a powerful and well-known concept in database marketing and were widely used to measure the value of customers based on their prior purchasing history [6]. The RFM concept had also been successfully integrated into the mining process in the past few years [7]–[9], giving rise to the idea that the RFM concept also can be applied to predict sales based on prior sales history.

In previous studies, the sales volume could be predicted by using a variety of methods, such as the neural network [10], [11], times series [11]–[13], support vector machines [14], [15] or machine learning [16], [17].

Meanwhile, previous studies related to modified RFM models such as RFM-CLV (RFM Customer Lifetime Value) which were proposed for customer segmentation [8], [18], RFM/P (RFM/Product) which estimates the value of customers per product and then estimates the total customer value [19], wRFM (weighted RFM) for customer segmentation [20], LRFM (Length RFM) which adopted self-organizing maps (SOM) techniques for segmentation of pediatric patients [21], and RFMT (RFM Time) is used to classify internet catering customer behavior [22].

These studies become the basic reference that the concept of RFM (modified RFM) can be adopted for sales forecasting that involves external factors that come from decision-makers.

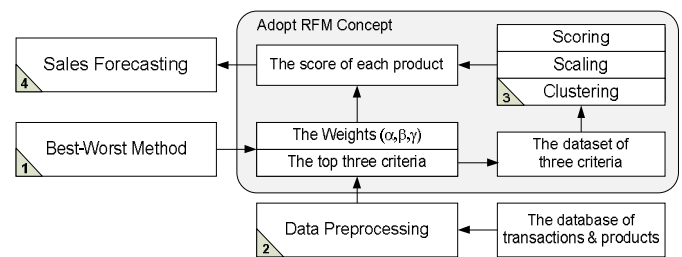
Therefore, this research will initiate the development of a novel sales forecasting method that adopts the RFM concept by modifying RFM criteria using BWM (one of the MCDM methods), then clustering transaction databases using data-mining method (such as k-Means) so that they can be used to sales forecasting.

The benefits that can be obtained from this research are: 1) can be an alternative method for sales forecasting, which plays a major role in business planning, so that management can increase market penetration capacity, optimize service to consumers, streamline inventory, and predict benefits the company's economy in the future; 2) can contribute to the development of sales forecasting method by innovating new approaches.

### 3. METHODOLOGY

The sales forecasting method proposed in this study will integrate the concept of recency-frequency-monetary (RFM) which is popularly used in data-mining projects to analyze customer value/behavior, data-mining clustering method (kMeans), and best-worst method (BWM) which is a method for decision-making.

The proposed method consists of five phases as illustrated in Figure 1.



**Figure 1:** The development phases of the proposed method.

**Phase 1.** Gathering the opinions of experts to determine the top three criteria (adopting the RFM concept) that most influence sales volume, then calculate the weight of the three criteria ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) using the best-worst method (BWM);

**Phase 2.** Data pre-processing such as sampling, transformation, denoising, and normalization will be carried out after transaction data and product data are imported from the case company database;

**Phase 3.** Data clustering uses data mining methods (such as the k-Means method) to determine the optimal number of clusters. Then scoring data for each criterion after the segmentation table (scaling) is obtained. The segmentation table which contains the top three criteria (outputs from Phase 1) will be the basis for scoring the database for each product;

**Phase 4.** After the final score of each product is obtained, the sales volume of each product for the coming month can be forecasted based on the multiplication of the final product score with last month's sales volume.

The output of each phase becomes the input of the next phase(s). After that, sales forecasting accuracy will be tested by comparing sales forecasting results with real sales data.

#### 3.1 Best-Worst Method

Multi-criteria decision-making (MCDM) is a very important branch of decision-making theory. In this research, a novel multi-criteria decision-making method called Best Worst Method (BWM) will be applied to determine the criteria affecting sales volume and its weights.

BWM is the latest multi-criteria decision-making (MCDM) method was proposed by Rezaei in 2015, which can obtain the weights of criteria and alternatives that compares the best criterion to the other criteria (alternatives), and all the other criteria to the worst criterion concerning different criteria based on pairwise comparisons with the need of less compared data [23].

The BWM is comprised of five steps that can be used to derive the weights of the criteria [23].

**Step 1:** Determine a set of decision criteria

Suppose there are  $n$  decision criteria ( $c_1, c_2, \dots, c_n$ ), which is very important for reasonably evaluating alternatives.

**Step 2:** Selecting the best and the worst criteria

In this step, the decision-maker selects the best (most influential) criterion and the worst (least influential) criterion among all the criteria identified in Step 1 from their perspective. Here only the criteria are selected and not the values of the criteria. The best criterion is represented as  $c_B$ , and the worst criterion is represented as  $c_W$ .

**Step 3:** Establish the preference rating for the best criterion over all the other criteria.

A number from 1 to 9 (1: equally important, 9: extremely more important) is used to indicate this value. The resulting a “Best-to-Others” vector as follows:  $\mathbf{a}_B = (a_{B1}, a_{B2}, \dots, a_{Bn})$ , where  $a_{Bj}$  represents the rating of the best criterion B over any other criteria  $j$  ( $j=1,2, \dots, n$ ), and  $a_{BB} = 1$ . The consensus of various experts is taken for the finalization of preference ratings.

**Step 4:** Establish the preference rating of all other criteria over the worst criterion.

A number between 1 and 9 is assigned in this case as well. The resulting an “Others-to-Worst” vector as follows:  $\mathbf{a}_W = (a_{1W}, a_{2W}, \dots, a_{nW})^T$ , where  $a_{jW}$  represents the rating of any criteria  $j$  over the worst criterion  $w$ , and  $a_{wW} = 1$ . Similarly, the final value can be arrived by consensus of all the experts involved in decision making.

**Step 5:** Calculating the optimal weights of all the criteria ( $w_1^*, w_2^*, \dots, w_n^*$ )

The objective is to calculate the weights of criteria so that the maximum absolute differences for all  $j$  are minimized of the following set  $\{|w_B - a_{Bj}w_j|, |w_j - a_{Bjw}w_w|\}$ , that can be formulated to the following min-max model:

$$\min \max \left\{ \left| \frac{w_B}{w_j} - a_{Bj} \right|, \left| \frac{w_j}{w_w} - a_{Bjw} \right| \right\}$$

Equation (1) can be transferred to the following linear programming model:

$$\begin{aligned} &\min \xi \\ &\text{s.t.} \\ &\left| \frac{w_B}{w_j} - a_{Bj} \right| \leq \xi, \text{ for all } j \\ &\left| \frac{w_j}{w_w} - a_{Bjw} \right| \leq \xi, \text{ for all } j \\ &\sum_j w_j = 1 \\ &w_j \geq 0, \text{ for all } j \\ &j = 1, 2, \dots, n \end{aligned}$$

Multiplying the first set of the constraints of (2) for any value of  $\xi$  by  $w_j$  and the second set of constraints by  $w_w$ , it can be seen that the solution space of (2) is an intersection of  $4n-5$  linear constraints (obtained from  $2(2n-3)$  of comparison constraint and 1 constraint for the number of weights) thus, provided that the solution is not empty enough. Solving (2), the optimal weights of the criteria ( $w_1^*, w_2^*, \dots, w_n^*$ ), and  $\xi^*$  are obtained.

After finding the final result, the consistency ratio (CR) of pairwise comparisons should be calculated because the consistency ratio is an important indicator for checking the consistency level of pairwise comparison. The consistency ratio of BWM can be calculated by using (3) as follows:

$$CR = \frac{\xi^*}{CI}$$

where CI (consistency index) is the maximum values of  $\xi$  for different values of  $a_{BW}$  is showed in Table 1,  $\xi^*$  is the solution of problem (2), and  $CR \in [0,1]$ . CR values close to 0 showed more consistency, while CR values close to 1 showed less consistency. From (3) can be resumed that the smaller the  $\xi^*$ , the smaller the CR, and the more consistent the vector is.

**Table 1:** Consistency Index (CI) [23]

$a_{BW}$	1	2	3	4	5	6	7	8	9
CI (max $\xi$ )	0.00	0.44	1.00	1.63	2.30	3.00	3.73	4.47	5.23

In this study, the Best-Worst Method will be used to determine the criteria and weight.

**3.2 Data Preprocessing**

After the top three criteria are generated from the BWM, the database fields of the case company associated with these three criteria will be selected, and then the data will be collected for preprocessing data.

Preprocessing data is needed to make knowledge discovery easier and more accurate. Preprocessing data consists of the data preparation process and data normalization. In the data preparation process, incomplete data is deleted from the dataset. Furthermore, the normalization process will be carried out to obtain a good data structure with the appropriate format, so the min-max approach can be applied in this research because all values of criteria are within the range between 0 and 1.

**3.3 The Concept of Recency-Frequency-Monetary**

RFM (recency, frequency and monetary) model is a behavior-based model used to analyze the behavior of a customer and then make predictions based on the behavior in

the database [24], [25]. In other words, RFM is a simple model for quantifying customer behavior.

The three behavioral attributes: Recency, Frequency, and Monetary are simple attributes that can be easily computed for any database according to the purchase history, and are easy to understand, also very powerful in predictive ability [25].

Recency is defined as the time length since the customer's last purchase; frequency is the number of purchases during the same period; and monetary is the total of money spent on all purchases given the same period.

This research will be used three criteria and their weight because it refers to the concept of RFM. The three criteria with the greatest weight ( $\alpha$ ,  $\beta$ , and  $\gamma$ ) are the results of evaluation using BWM and will be assessed based on the table of scaling of criteria.

#### A. Clustering

Generally, in the RFM concept, the dataset for each criterion will be segmented into 5 equal parts or each part is worth 20% of the entire dataset. But to optimize the dataset segmentation in this study, the number of clusters will be determined by applying a clustering method (k-means method) [3].

#### B. Scaling of Criteria

This step will define the interval for the value of each selected criterion. In short, the result of this process is the determination of the interval of values for each selected criterion (criteria scale table) that will be used in the criteria scoring process.

#### C. Criteria Scoring

Criteria scoring aims to project customer value or convert customer value into numbers so that it can be used in calculations. The projected score of each criterion for each product in this study will be tabulated on a scale table.

Scores of each product are calculated by referring to the RFM score calculation as follows [8]:

$$\begin{aligned} \text{Product score} &= \text{score} * \text{weight of criterion 1} + \\ &\quad \text{score} * \text{weight of criterion 2} + \\ &\quad \text{score} * \text{weight of criterion 3} \end{aligned}$$

In the form of a simple formula, the product score can be defined as (4) where  $w_1$ ,  $w_2$ , and  $w_3$  denote the weights of each criterion, and  $s_1$ ,  $s_2$ ,  $s_3$  denote the scores of each criterion derived from the scale table. The product score will be a multiplier factor in forecasting its sales volume [8].

$$\text{Product score} = s_1 .w_1 + s_2 .w_2 + s_3 .w_3 \quad (4)$$

### 3.4 Sales Forecasting

Forecasting the sales volume of a product for the coming month can be obtained by multiplying this month's sales volume with this month's product score. Thus, the sales volume of a product this month will be the basis for forecasting product sales volume in the coming month by involving external factors.

## 4. DISCUSSION

In this research paper, several phases of the method illustrated in Figure 1 can be further developed for subsequent research, as in phase 1, other decision-making methods can be used, such as Fuzzy ANP. In phase 2, data pre-processing methods such as Principal Component Analysis can be applied, and in phase 3 other clustering methods such as Particle Swarm Optimization can be used. A comparison of trial results with various combinations of various development methods will produce the best sales forecasting method.

Among all the explanations for this finding is that the more combinations of methods tested, the better the sales forecasting method produced. However, this paper is still a concept of a new sales forecasting method. Future research must be carried out by gathering data relevant to the process of analysis and testing of the proposed method.

## 5. CONCLUSION

This research has resulted in a new sales forecasting method that integrates decision-making method, data mining clustering method, and RFM concept. Literature studies and phases in developing the sales forecasting method have been explained in detail in this paper. The results of this study are still in the form of a concept that can be further developed and tested using a real company database. The smaller the error rate produced, the higher the accuracy of the forecasting method. We are optimistic that this new sales forecasting method can be an alternative method for company management to increase market penetration capacity, optimize services to consumers, inventory efficiency, and predict the company's economic benefits in the future.

These findings can provide benefits for further research such as the analysis and evaluation related to the proposed sales forecast method, developing sales forecasting applications such as [26] to facilitate the sales forecasting process, and reference the latest literature reviews in the area of sales forecasting.

## ACKNOWLEDGMENT

The authors would like to thank the Ministry of Research and Technology/National Research and Innovation Agency, Republic of Indonesia (Kemenristek/BRIN RI) for financially supporting this research.

## REFERENCES

1. M. Y. Ma'arif, "A Mediating Effect on ERP KM Model for the Performance of Oil and Gas Sector in Klang Valley: A Preliminary Study," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1.4, pp. 463–468, Sep. 2019.  
<https://doi.org/10.30534/ijatcse/2019/7381.42019>
2. M. Alif Basar, "A Preliminary Study on Mechanism and Measurement of Trustworthiness on a Crowdsourcing Platform," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1.3, pp. 409–415, Aug. 2019.  
<https://doi.org/10.30534/ijatcse/2019/7181.32019>
3. R. Gustriansyah, N. Suhandi, and F. Antony, "Clustering optimization in RFM analysis based on k-means," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 18, no. 1, pp. 470–477, 2020.
4. N. Suhandi, Terttiaavini, and R. Gustriansyah, "Decision support system for subsidized housing selection based on best-worst method and simple additive weighting," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 3, pp. 3384–3389, 2020.  
<https://doi.org/10.30534/ijatcse/2020/138932020>
5. J. Rezaei, "Best-worst multi-criteria decision-making method: Some properties and a linear model," *Omega*, vol. 64, pp. 126–130, Oct. 2016.
6. H. A. Mahdiraji, E. Kazimieras Zavadskas, A. Kazeminiya, and A. Abbasi Kamardi, "Marketing strategies evaluation based on big data analysis: a CLUSTERING-MCDM approach," *Econ. Res. Istraživanja*, vol. 32, no. 1, pp. 2882–2898, Jan. 2019.
7. J. T. Wei, S.-Y. Lin, Y.-Z. Yang, and H.-H. Wu, "Applying Data Mining and RFM Model to Analyze Customers' Values of a Veterinary Hospital," in *2016 International Symposium on Computer, Consumer and Control (IS3C)*, 2016, pp. 481–484.
8. S. Monalisa, P. Nadya, and R. Novita, "Analysis for Customer Lifetime Value Categorization with RFM Model," *Procedia Comput. Sci.*, vol. 161, pp. 834–840, 2019.  
<https://doi.org/10.1016/j.procs.2019.11.190>
9. E. Gharavi and M. J. Tarokh, "Predicting customers' future demand using data mining analysis: A case study of wireless communication customer," in *The 5th Conference on Information and Knowledge Technology*, 2013, pp. 338–343.
10. N. Caglayan, S. I. Satoglu, and E. N. Kapukaya, "Sales Forecasting by Artificial Neural Networks for the Apparel Retail Chain Stores," in *Advances in Intelligent Systems and Computing*, 2020, pp. 451–456.
11. V. Borysenko, G. Kondratenko, I. Sidenko, and Y. Kondratenko, "Intelligent forecasting in multi-criteria decision-making," in *CEUR Workshop Proceedings*, 2020, vol. 2608, pp. 966–979.
12. R. Gustriansyah, N. Suhandi, F. Antony, and A. Sanmorino, "Single Exponential Smoothing Method to Predict Sales Multiple Products," *J. Phys. Conf. Ser.*, vol. 1175, no. 1, pp. 1–7, 2019.
13. J. P. Karmy and S. Maldonado, "Hierarchical time series forecasting via Support Vector Regression in the European Travel Retail Industry," *Expert Syst. Appl.*, vol. 137, pp. 59–73, Dec. 2019.
14. S. Zhang and D. Liang, "SVM and TOPSIS analysis of demand forecast management model based on discretion correction factor," *Int. J. Electr. Eng. Educ.*, pp. 1–12, Dec. 2019.
15. Y. Zhu, Y. Zhao, J. Zhang, N. Geng, and D. Huang, "Spring onion seed demand forecasting using a hybrid Holt-Winters and support vector machine model," *PLoS One*, vol. 14, no. 7, pp. 1–18, Jul. 2019.  
<https://doi.org/10.1371/journal.pone.0219889>
16. A. Martínez, C. Schmuck, S. Pereverzyev, C. Pirker, and M. Haltmeier, "A machine learning framework for customer purchase prediction in the non-contractual setting," *Eur. J. Oper. Res.*, vol. 281, no. 3, pp. 588–596, Mar. 2020.
17. S. Cheriyan, S. Ibrahim, S. Mohanan, and S. Treesa, "Intelligent Sales Prediction Using Machine Learning Techniques," in *2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, 2018, pp. 53–58.
18. F. Yoseph and M. Heikkila, "Segmenting Retail Customers with an Enhanced RFM and a Hybrid Regression/Clustering Method," in *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)*, 2018, pp. 108–116.
19. R. Heldt, C. S. Silveira, and F. B. Luce, "Predicting customer value per product: From RFM to RFM/P," *J. Bus. Res.*, Nov. 2019.
20. A. Z. Ravasan and T. Mansouri, "A Fuzzy ANP Based Weighted RFM Model for Customer Segmentation in Auto Insurance Sector," in *Intelligent Systems, IGI Global*, 2018, pp. 1050–1067.
21. J.-T. Wei, S.-Y. Lin, C.-C. Weng, and H.-H. Wu, "A case study of applying LRFM model in market segmentation of a children's dental clinic," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 5529–5533, Apr. 2012.  
<https://doi.org/10.1016/j.eswa.2011.11.066>
22. X. Hu, Z. Shi, Y. Yang, and L. Chen, "Classification Method of Internet Catering Customer Based on Improved RFM Model and Cluster Analysis," in *2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, 2020, pp. 28–31.
23. J. Rezaei, "Best-Worst Multi-Criteria Decision-Making Method," *Omega - Int. J. Manag. Sci.*, vol. 53, pp. 49–57, 2015.

24. A. M. Hughes, “**Boosting response with RFM. Mark,**” *Tools*, vol. 5, pp. 4–10, 1994.
25. I.-C. Yeh, K.-J. Yang, and T.-M. Ting, “**Knowledge discovery on RFM model using Bernoulli sequence,**” *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5866–5871, Apr. 2009.  
<https://doi.org/10.1016/j.eswa.2008.07.018>
26. R. Gustriansyah, N. Suhandi, and F. Antony, “**The Design of UML-Based Sales Forecasting Application,**” *Int. J. Recent Technol. Eng.*, vol. 7, no. 6, pp. 1507–1511, 2019.