

# A novel approach to optimizing customer profiles in relation to business metrics

Marischa Elveny<sup>1</sup>, Mahyuddin K. M. Nasution<sup>1,3</sup>, Muhammad Zarlis<sup>2</sup>, Syahril Efendi<sup>1</sup>,  
Rahmad B. Y. Syah<sup>3</sup>

<sup>1</sup>Data Science of Artificial Intelligence, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia

<sup>2</sup>Master of Information System Management, Universitas Bina Nusantara, Jakarta, Indonesia

<sup>3</sup>Excellent Centre of Innovations and New Science, Faculty of Engineering, Universitas Medan Area, Medan, Indonesia

## Article Info

### Article history:

Received Nov 28, 2023

Revised Feb 12, 2023

Accepted Mar 10, 2023

### Keywords:

Business metrics

Customer profile

Multivariate adaptive regression

spline

Prediction

Robust

## ABSTRACT

Business is very closely related to customers. Each user owns the data, and the data is used to identify cross-selling opportunities for each customer. For example, the type of product or service purchased, the frequency of purchases, geographic location, and so on. By doing so, you can gain the ability to manage and analyze customer data, allowing you to create new opportunities in industries that were previously difficult to enter. The purpose of optimizing user profiles is to determine minimum or maximum business value and improve efficiency by determining user needs. In this study, multivariate adaptive regression spline (MARS) is a statistical model used to explain the relationship between the response variable and the predictor variable. Robust is used to find variable relationships to make predictions. To improve classification performance, the model is validated using a confusion matrix. The results show an accuracy value of 84.5%, with better time management (period management) reflected in the number of hours spent by merchants as well as discounts during that time period, which has a significant impact on any business. In addition, the distance between customers and merchants is also important, as customers prefer merchants who are closer to them to save time and transportation costs.

*This is an open access article under the [CC BY-SA](#) license.*



## Corresponding Author:

Marischa Elveny

Data Science of Artificial Intelligence, Faculty of Computer Science and Information Technology

Universitas Sumatera Utara

Padang Bulan 20155 USU, Medan, Indonesia

Email: marischaelveny@usu.ac.id

## 1. INTRODUCTION

People whose businesses are inextricably linked to the customer. Every customer has data, and the existence of this data serves a purpose in policy implementation [1]. Understanding customer data is critical for running a business. A customer profile is a feature or variable that identifies a customer. They can be either psychographic or demographic in nature [2]. Some examples include age, gender, occupation, place of residence, social class, and purchasing behavior. Profiles assist businesses in learning more about their customers, such as who they are, what they purchase, and where they purchase [3].

Companies can plan their marketing mix with this knowledge in mind. Businesses must create profiles in order to identify and measure the habits and characteristics of consumers in the target market. The majority of marketing strategies fail because they fail to target anyone while attempting to appeal to everyone. As a result, the company sells customers unnecessary products. Profile knowledge is used to segment the market. Businesses

can use segmentation to match their products and services to the needs of their customers, thereby enhancing the success of their marketing strategies [4].

Business intelligence (BI) is defined as a collection of techniques and tools for gathering and transforming raw data into meaningful information for business analysis [5]. The digital era has truly changed society, as evidenced by the presentation of e-metrics business designs that make it easier for users (customers) to behave electronically and provide information that can be applied [6]. Similarly, issues frequently arise, particularly when attempting to obtain information about how to predict behavior, particularly when working with large amounts of data. In the context of BI, the most important feature is the condition of the amount of big data. To produce predictive patterns of customer behavior, the concept of data mining must be investigated [7]. Because there is so much data to manage, predictions will be made using the concept of strong non-parametric regression.

The method is used to track the origin of information or predict social behavior by analyzing customer profiles based on degree of behavioral similarity [8]. Prediction optimization refers to the process of systematically predicting what will happen based on data from the past and present. This connection is used to build models and forecast future events [9]. Some of the regression assumptions are violated by the nonparametric regression model, resulting in a less accurate predictive value. Multivariate adaptive regression spline (MARS) is used in linear regression analysis to handle outlier data and to model the relationship between each multivariable that appears [10]. Robust is a method for solving optimization problems with uncertain data that are only known to belong to one of several uncertainty sets in the presence of outliers [11]. Robust's goal is to find optimal or near-optimal solutions for any given realization of uncertain scenarios [12]. The goal of this research is to develop an optimization model for predicting customer profiles based on multiple predictor variables that can be used by competitive organizations to make decisions.

## 2. METHOD

Based on registered customers, mobile wallets generate data collections [13]. During the processing stage, several stages were completed, including data processing using the multivariate adaptive regression spline (MARS) method is used to process data [14]. MARS identified customer behavior as a determining factor in reducing data deviation. Robust detects and resolves outliers using deviations before optimizing the data. The prediction results will use a new model to perform performance on the data, and model validation will be done using the Confusion matrix [15]. During the testing process, the Confusion Matrix is used to assess the appropriateness of calculation accuracy [16]. Further tests were conducted with various data distribution compositions for training and testing, namely 55%:45%, 65%:35%, 80%:20%, and 90%:10%. This is done to obtain a more accurate model comparison. Figure 1 depicts the research steps.

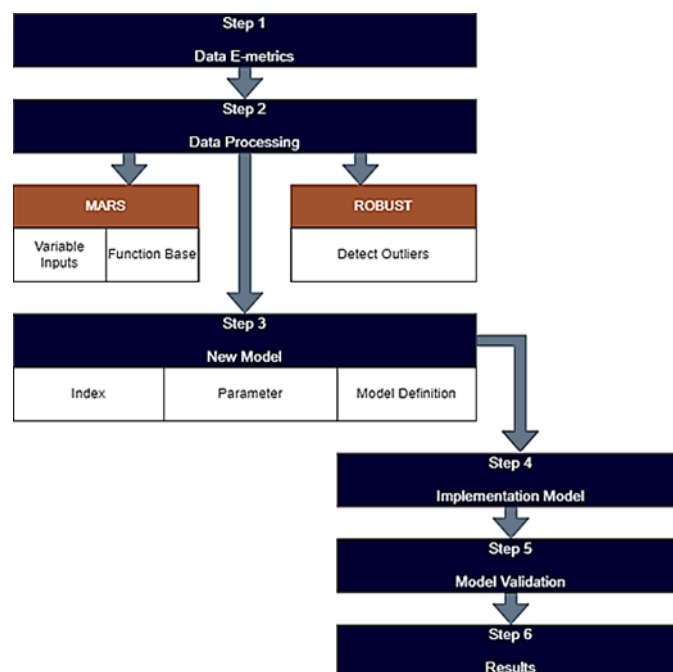


Figure 1. State of the art

## 2.1. E-Metrics

E-metrics refers to customer behavior that is tracked electronically (e-customer behavior) [17]. As evidenced by the ability to present everything electronically, provide the technological era has brought about significant changes in society by providing convenience about how a customer behaves electronically and providing information that can be applied [18]. The number and frequency of merchants with Variants vary according to the type of business actor, especially in the digital space [19]. Each merchant's user Financial Technology is used in e-metrics to track behavior activities. Figure 2 depicts the e-metrics ecosystem.

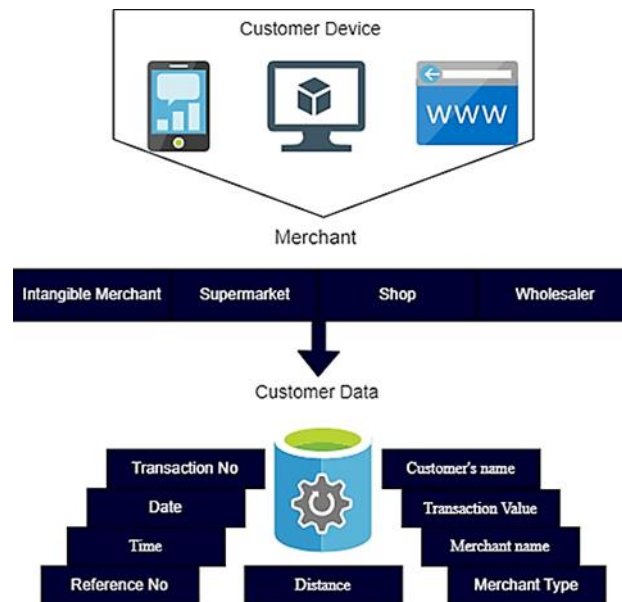


Figure 2. Ecosystem e-metrics

## 2.2. Customer lifecycle

The customer lifecycle is a method of managing the various stages of the customer interaction process with the system in detail and comprehensive [20], namely the customer experience is stored and managed from start to finish, can be seen in Figure 3. This is required as a foundation for the system to learn about customer habits. Figure 3 explains that customers register, then make requests for the desired product, and then the data is investigated. Data investigation is intended to sort the data and determine whether or not the data is appropriate. Customers who pass the investigation can place product orders, transactions, and payments. Furthermore, the system will record all customer habits as a type of variable in data processing.

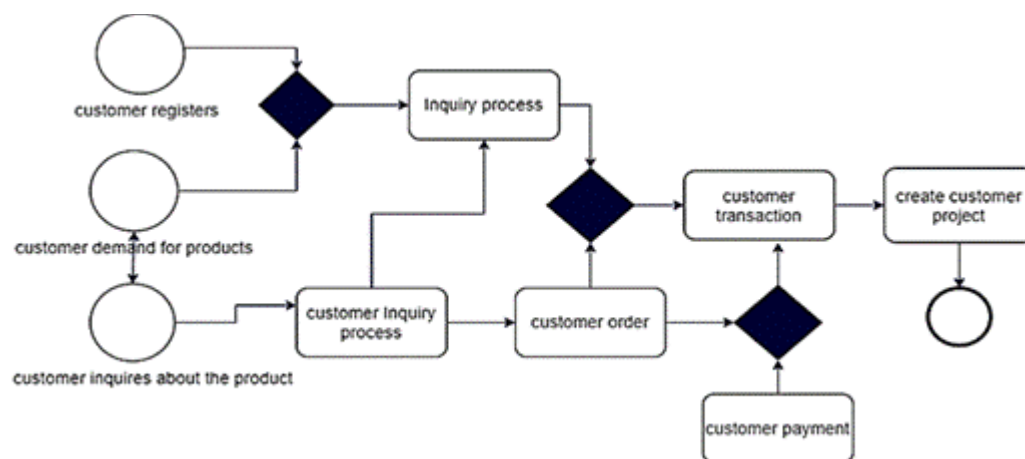


Figure 3 Customer lifecycle

### 2.3. Multivariate adaptive regression spline (MARS)

Multivariate adaptive regression splines (MARS) is a high-dimensional data modeling technique. The model takes the form of an extension in the basic function of the product spline, with the data determining the number of basic functions and parameters associated with each function. This procedure is motivated by recursive partitioning and has the capability of capturing high-order interactions or involving the greatest number of interactions in a few variables and generating a continuous model. Furthermore, when identifying the additive contribution associated with multivariable interactions, the model can be represented in a discrete form. In this study, the MARS method was used to analyze the data [21]. To get the best model, follow these steps:

- Ascertain that each variable (response and predictor variables) has the same value scale. MARS employs a piecewise linear expansion of the form

$$[+(a - P)]_+, [-(a - P)]_+, \quad (1)$$

- As a first step in determining the relationship pattern between these variables, a graphic plot is created each predictor variable and each response variable. For each predictor variable a general model of the relationship between the predictor and response variable attempts to construct a reflected pair  $a_j$  ( $j = 1, 2, p$ ) with  $p$ -dimensional knots  $\tau_i = (\tau_{i,1}, \tau_{i,2}, \dots, \tau_{i,p})^T$  at  $a_i = (a_{i,1}, a_{i,2}, \dots, a_{i,p})^T$  or just next to each data data vectors  $a_i = (a_{i,1}, a_{i,2}, \dots, a_{i,p})^T$  ( $i = 1, 2, \dots, N$ ).
- Determine the number of available basis functions (BF), which should be two to four times the number of predictor variables. Due to the use of 9 predictor variables, the maximum number of BFs in this study were 14, 21, and 28.

$$\wp : \{(a_j - P)_+, (P - a_j)_+ \mid P \in \{a_{1,j}, a_{2,j}, \dots, a_{i,j}\}, J \in \{1, 2, \dots, p\}\}, \quad (2)$$

- Determine the maximum number of interactions that can occur in this study (MI). are numbered one, two, and three Because more than three interactions will result in a model interpretation that is extremely complex.
- Because there is no basis or fixed boundary for determining the minimum number of observations between nodes, use trial and error.
- Run parameter tests on the best model using the minimum Generalized cross-validation (GCV) value criteria from all possible combinations of basis functions (BF), maximum interaction (MI), and minimum observation (MO) values. Generalized cross-validation (GCV) can be used to indicate that the MARS model does not fit.

$$GCV := \frac{\frac{1}{X} \sum_{i=1}^X (y_i - \hat{f}_\alpha(a_i))^2}{(1 - \bar{c}(\alpha)/X)^2} \quad (3)$$

- Concurrently or partially test the MARS model's significance with the regression coefficient test (F test) (t test).
- Explain the MARS model and the variables that affect it.

### 2.4. Function base (BF)

To estimate the response variable in this study, 9 predictor variables were used. There are one, two, or three interaction variables (MI) that can be determined. Because a model with more than three interactions will result in a very complex interpretation. We chose minimum observations (MO) of 0, 1, 2, 3, 4, and 5 [22]. This study makes use of a nonparametric model, with (4).

$$h = f(a) + \varepsilon \quad (4)$$

Where  $h$  denotes the response variable and  $x$  denotes a constant.  $(a_1, a_1, \dots, a_p)^T$  is the predictor variable, which is a stochastic additive component with a finite variance and a zero mean. Following that, the data is analyzed to determine the optimal number of subregions and functions. optimally in accordance with the realization of each subregion [23]. This study can solve the optimization problem using node values as data points, but because the values are so close to each variable, the best Base Function Model is required to find competitive predictions, as shown in (4).

$$h = -0.5464(a_1) + 0.06474(a_2) + 0.001125(a_3) - 0.04673(a_4) + 0.04637(a_5)$$

$$+0.06778(a_6) + \epsilon$$

Table 1 is generated based on (4). Using Table 1, Find the maximum basis function, which should be two to four times the number of predictor variables used [24]. Meanwhile, the smallest observations are 0 (one), 1, 3, 5, 7, and 9.

Table 1. Function base value

Parameter	Estimate	S.E.	T-Ratio	P-Value
Constant	11658.3379	1431.64758	15.36475	0.00000
BF 1	772654.883	4482.37586	15.47584	0.00000
BF 3	-88753.7463	4268.25465	-19.2537	0.00000
BF 5	-66538.8763	5546.11530	-15.46586	0.00000
BF 7	167676E+05	26489.51622	13.73948	0.00000
BF 9	-63437.3246	55637.23844	-7.37596	0.00000
BF 14	-0.00001	0.00001	-3.28384	0.00007
BF 21	-0.00001	0.00001	-4.476593	0.00002
BF 28	234.5567	74.374428	3.36486	0.00256

$$BF14 = \max(0, h_4 - 9,019e+011) * BF2;$$

$$BF21 = \max(0, 3648e+011 - h_4) * BF2;$$

$$BF28 = \max(0, h_2 - 6785) * BF2;$$

$$h = 2447,8 + 7547,7 * BF1 - 63785,2 * BF3 - 6484 * BF5 + 63748 * BF7 - 1435,8 * BF9 - 3,6474 - 007 * BF14 - 7,637-006 * BF1 + 766,909 * BF21;$$

$$\text{Piecewise Linear GCV} = 537433 + 0,9 = 537,433$$

Table 2 displays the sensitivity and specificity values generated by the model's interpretation. Table 2 is generated based on (5). Specificity is defined as the proportion of correctly identified class 0 that is truly negative [25]. The proportion of truly positive, i.e., the proportion of class 1 that is correctly identified, is measured by sensitivity. The following is the specificity and sensitivity equation:

$$\begin{aligned} \text{Specificity (\%)} &= \frac{f_{00}}{f_{10} + f_{11}} \times 100 \\ \text{Sensitivity (\%)} &= \frac{f_{11}}{f_{10} + f_{11}} \times 100 \end{aligned} \quad (5)$$

Table 2. Sensitivity dan specificity

Score Range	If Score	Sensitivity	Specificity
1.00000	> 1.00000	0.00000	1.00000
0.99999	> 0.99999	0.26563	1.00000
0.00000	> 0.00000	1.00000	1.00000
0.00000	> 0.00000	1.00000	0.73373
0.00000	> 0.00000	1.00000	0.73538
0.00000	> 0.00000	1.00000	0.33436
0.00000	> 0.00000	1.00000	0.16354
0.00000	> 0.00000	1.00000	0.10973
0.00000	> -0.00000	1.00000	0.06353
0.00000	> -0.00000	1.00000	0.05354

There is an error in the training data when calculating the sensitivity of the transaction value to the merchant, as shown in Table 2, and the results of the error calculation are shown in Table 3. Table 3 In forecasting, the mean squared error (MSE) is displayed, double-check the estimation of the error value. A low MSE is one that is close to zero, indicates that the forecasting results are consistent with actual data and can be used to forecast the future [26]. The MSE value, on the other hand, is close to zero. In mean absolute deviation (MAD), it's used to figure out the average absolute or absolute error. MAD determines forecast accuracy by averaging estimated errors (absolute value of each error). The Root Mean Square Error measures the difference in the predicted value of a model as an estimate of the observed value.

Table 3. Results of the MARS data training error calculation data model

Percentile	N	MSE	MAD	RMSE	MRAD
99.76%	710	0.929	0.3409	0.9640	0.24077
99.30%	706	0.509	0.2952	0.7138	0.23758
98.59%	701	0.332	0.2615	0.5766	0.23419
96.48%	686	0.173	0.2102	0.4163	0.20633
92.97%	661	0.1192	0.1706	0.3453	0.14726
85.94%	611	0.5146	0.1050	0.2268	0.08226

## 2.5. Robust

According to (6), where  $R$  represents the random response variable and  $R = (R_1, R_2, \dots, R_p)^T$  is the random vector predictor variable, and all of them contain "noise." For each input  $R_j$ : [27] We set the  $U1$  and  $U2$  uncertainties in the data model for robust as polyhedrals for input and output to avoid increasing the complexity of the optimization problem.  $U1$  and  $U2$  are determined by the set [28]. The term "robust" can be defined in (7):

$$R_j = R + \mu_j (j = 1, 2, \dots, p) \quad (6)$$

$$\min_{\alpha} \max_{\substack{W \in U1 \\ z \in U2}} \|z - w\epsilon\|_2^2 + \phi \|L\alpha\|_2^2 \quad (7)$$

$U1$  is a polytype with  $2^{N-M_{max}}$  with an angle  $W^1, W^2, \dots, W^{2N-M_{max}}$  based on the MARS error calculation results, look for outliers from uncertainty. To begin, make sure that the input and output variables are distributed evenly [29]. The model was then built using Salford MARS Version 3. The noise (uncertainty) was then incorporated into the real input data in order to apply a robust optimization technique to the MARS model [30]. General algebraic modelling system (GAMS) Studio37 is a powerful optimization solution.

## 3. RESULT AND DISCUSSION

### 3.1. Contribution model

The forward stepwise part of the MARS algorithm determines the node points between data points to obtain BFs. Increasing the number of endpoints results in an increase in the number of data points. As a result, complications arise. As a result, using clustering theory, statistics, and optimal theory, it is possible to determine the node points based on the data set. As a result, the model is built with

- Consider increasing the flexibility index in order to improve customer habits.
- Using learning phenomena in the manufacturing process, such as discounts at specific times. Process enhancements that result in an increase in customers over a specific time period.
- The goal of taking into account features is to maximize flexibility while minimizing risk.

### 3.2. Model definition

- M1: Min transaction value, on (8).

$$\begin{aligned} & \sum_s \sum_t (htpp_{st} \times mp_{st}) \sum_p \sum_t (btp_{pt} \times mp_{pt}) \\ & + \sum_d \sum_t (pt_{dt} \times dp_{dt}) \sum_s \sum_t (btp_{st} \times kp_{st}) + \sum_p \sum_t (btp_{pt} \times km_{pt}) \\ & + \sum_d \sum_t (fud_{dt} \times ksp_{dt}) \sum_i \sum_p \sum_t (bpm_{ipt} \times pp_{ipt}) + \sum_p \sum_t (bpp_{idt} \times ppp_{idt}) \end{aligned} \quad (8)$$

- M2: Min time management (period), on (9).

$$\sum_i \sum_d \sum_c \sum_l \sum_t (jppt_{idct} \times lwp_{calt} \times jmp_{acl}) \quad (9)$$

- M3: Max flexibility/optimization, on (10).

$$\frac{\sum_t (DVF_t)}{|T|} \quad (10)$$

- M4: Min risk/distance, on (11).

$$\begin{aligned} & \sum_i \sum_s \sum_t (rp_{ist} \times mp_{st}) + \sum_i \sum_p \sum_t (rpp_{ipt} \times mm_{pt}) \\ & + \sum_i \sum_d \sum_t (rpm_{idt} \times mmp_{dt}) \end{aligned} \quad (11)$$

The (8) is a mathematical representation of the first objective function, which examines behavior through transactions. The model's second goal is to minimize the function for time or period management, taking into account the use of discounts at specific times as shown in (9). The third goal is to maximize the flexibility of all customer activities, beginning with the time, type of transaction, type of product, and transaction costs as shown in (10). The fourth objective function in (11) is intended to reduce risk based on distance between the customer and the destination merchant [31], [32].

### 3.3. Model implementation

From January to December, the number of products purchased is shown in Table 4. The table shows which products were purchased the most from January to June. While there was a decrease in sales from July to December. This can be used as a guideline when deciding whether to increase sales during these months. Table 5 explains that transactions per time are carried out at 07.00-22.30 WIB. Needed to see the habits of customers making transactions at certain hours. where this can be used as a reference by merchants in opening and closing their shops.

Table 4. Grouping by product

Month	Food	Drink
January	225	120
February	163	158
March	190	93
April	129	89
Mei	88	52
June	130	51
July	54	38
August	56	43
September	58	44
October	47	50
November	44	41
December	52	32

Table 5. Number of transactions per time

Hour	Food	Drink
07.00 WIB - 11.59 WIB	309	195
12.00 WIB - 16.59 WIB	457	278
17.00 WIB - 22.30 WIB	390	251

Figure 4 depicts the number of transactions per time division starting at 07.00-11:59 WIB, 12.00-16:59 WIB, and 17.00-22.30 WIB. Where this display shows the number of transactions at any given time. The highest number of transactions can be seen at 12:00-16.59 WIB.

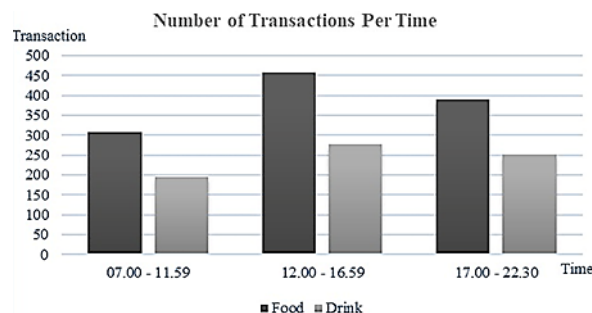


Figure 4. Number of transactions per time

### 3.4. Analysis of user habits based on indicators

Analyzing user behavior based on transaction and merchant indicators is accomplished by examining user habits when transacting with a merchant [33]. The study was conducted based on the estimated maximum limit. Looking at the number of variables that appear, there are 448.

The rupiah currency was used in this study. According to Figure 5, the number of transactions made by customers ranges from 0 to 120, and the total payment made ranges from 0 to 8,000 rupiah. Figure 5 shows that there were more than 80 transactions with prices ranging from 70,000 to 80,000 rupiah. It can be concluded that price is an important indicator in determining customer behavior patterns.

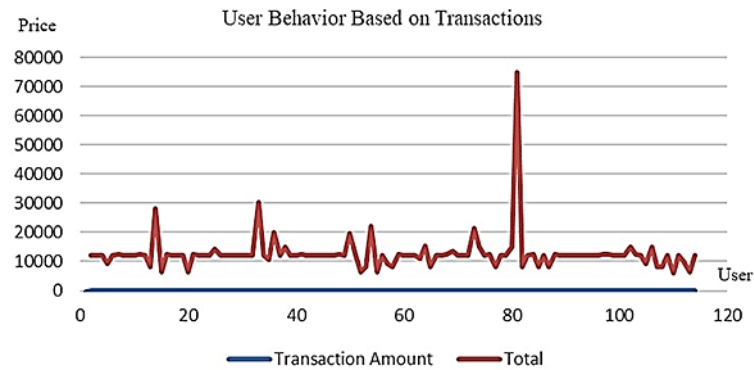


Figure 5. User behavior based on transactions

### 3.5. Results

After the model is trained, data performance testing is then carried out. The results of the analysis aim to find out how the comparison between testing data and training data affects the best results. Test data is done with a comparison of 65% test data and 35% training data. To validate the model developed in this study, the Confusion Matrix technique was used. By calculating accuracy, the confusion matrix is used to assess the suitability of the data testing process.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (12)$$

Information:

*TP* = True Positive, predicts the correct customer profile for transactions

*TN* = True Negative, prediction of customer profiles that do not transact

*FP* = False Positive, prediction of registered customer profile, but the fact is the transaction

*FN* = False Negative, prediction of registered customer profile, but in fact no transaction

The confusion matrix uses measurements based on precision, recall, F1-score, and accuracy [34] and the results show that the distribution of the training data is 485 and the test data is 898, with a division of 65%: 35%. Get the best results. With a total accuracy of 84.5%. The findings of this study can be seen in Table 6 which explains the findings of the analysis based on the four functional models implemented in the data.

Table 6. Optimal sum of each objective function with problem solving

Optimal Quantity	Function M1	Function M2	Function M3	Function M4
Small	10,673,920	31,887,535	49.96	9.13
Currently	31,887,554	22,754,838	84.24	8.15
Big	31,887,535	44,358,021	9.13	9.5

The results show that M3 and M4 functions have a higher level, where the value is close to 0. The M3 function expands the size in terms of time management (period), which is for example the length of merchant opening time, the existence of discounts at specific times, has a significant impact. in the first objective function, which can increase the number of transactions Meanwhile, the M4 function, i.e., the distance between the customer and the merchant, has a significant impact. Customers are looking for nearby merchants because of the efficiency in terms of time and transportation costs.



These findings can be used to help manage new businesses where innovation emerges from interactions between customers and merchants. This viewpoint is important because ongoing marketing innovations may require a rethink about competition and collaboration between business groups. can be seen in Figure 6 explains that customer and merchant relationships are based on flexibility, customer habits, processing time, discounts and rewards as well as flexibility in minimizing risk. As:

- Customer interface: how downstream customer relationships are structured and managed
- Costs and benefits of the financial model
- distribution among stakeholders in the business model

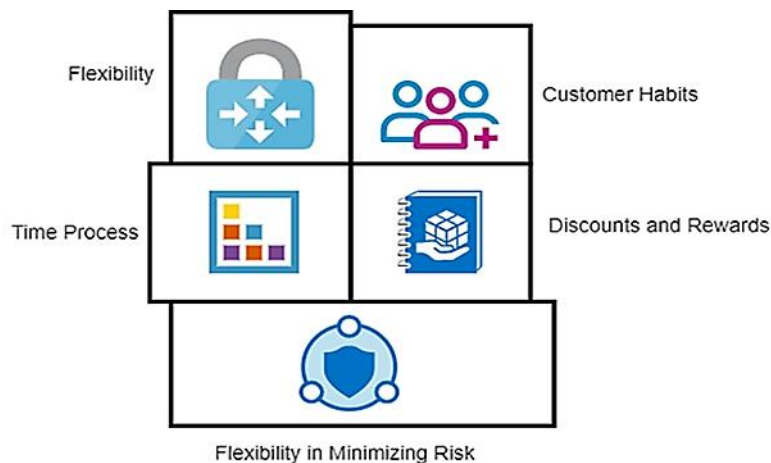


Figure 6. Analysis for business model innovation

#### 4. CONCLUSION

The study's findings were obtained through various studies of information poured into the knowledge base. Specifically, MARS data was generated using a function basis to find maximum values with upper, middle, and lower limits, as well as interactions that serve as a match between input variables until simplification is achieved. Robust can optimize outliers in data, resulting in new models. The new model was built by considering four main functions: transaction behavior, time or period management, maximizing the flexibility of all customer activities, starting from time, transaction type, product type, and transaction costs, as well as risk reduction based on the distance between the customer and the destination merchant. dividing the data 65:35% gives the best results. This is reflected in the total accuracy of 84.5%, where the model explains that improvements to time management (period), such as the length of merchant opening time or the availability of discounts at certain times, have a significant impact. Meanwhile, the distance between customers and merchants has a significant impact. Customers look for the nearest merchant because of the efficiency of time and transportation costs. We undertook this review to increase the conceptual clarity around what a business model and an innovation model mean. The main contribution of this paper is to pave the way for shared understanding and how to innovate for the development of research that will support academics and industry practitioners as well as business people in making decisions by increasing customer profiles.

#### ACKNOWLEDGEMENTS

Thanks to DRTPM for the research funding provided based on Decree Number 059/E5/PG.02.00/PT/2022 and Agreement/Contract Number 1279/UN5.2.3.1/PPM/KP-DRTPM/L/2022.




#### REFERENCES

- [1] P. Rikhardsson and O. Yigitbasioglu, "Business intelligence & analytics in management accounting research: Status and future focus," *International Journal of Accounting Information Systems*, vol. 29, pp. 37–58, Jun. 2018, doi: 10.1016/j.accinf.2018.03.001.
- [2] M. R. Llave, "A review of business intelligence and analytics in small and medium-sized enterprises," *International Journal of Business Intelligence Research*, vol. 10, no. 1, pp. 19–41, Jan. 2019, doi: 10.4018/IJBIR.2019010102.
- [3] M. Bardicchia, *Digital CRM: Strategies and Emerging Trends: Building Customer Relationship in the Digital Era Pdf*. 2020.
- [4] N. Svensson and E. K. Funck, "Management control in circular economy. Exploring and theorizing the adaptation of management control to circular business models," *Journal of Cleaner Production*, vol. 233, pp. 390–398, Oct. 2019, doi:




- 10.1016/j.jclepro.2019.06.089.
- [5] V. H. Trieu, "Getting value from Business Intelligence systems: A review and research agenda," *Decision Support Systems*, vol. 93, pp. 111–124, Jan. 2017, doi: 10.1016/j.dss.2016.09.019.
  - [6] N. U. Ain, G. Vaia, W. H. DeLone, and M. Waheed, "Two decades of research on business intelligence system adoption, utilization and success – A systematic literature review," *Decision Support Systems*, vol. 125, p. 113113, Oct. 2019, doi: 10.1016/j.dss.2019.113113.
  - [7] D. Arnott, F. Lizama, and Y. Song, "Patterns of business intelligence systems use in organizations," *Decision Support Systems*, vol. 97, pp. 58–68, May 2017, doi: 10.1016/j.dss.2017.03.005.
  - [8] N. A. El-Adaileh and S. Foster, "Successful business intelligence implementation: a systematic literature review," *Journal of Work-Applied Management*, vol. 11, no. 2, pp. 121–132, Sep. 2019, doi: 10.1108/JWAM-09-2019-0027.
  - [9] C. A. Tavera Romero, J. H. Ortiz, O. I. Khalaf, and A. R. Prado, "Business intelligence: business evolution after industry 4.0," *Sustainability (Switzerland)*, vol. 13, no. 18, p. 10026, Sep. 2021, doi: 10.3390/su131810026.
  - [10] Y. Kahane, "New Metrics for a New Economy: The B2T by 2020 Project," in *Palgrave Studies in Sustainable Business in Association with Future Earth*, Springer International Publishing, 2019, pp. 101–113. doi: 10.1007/978-3-030-14199-8\_6.
  - [11] S. Rakshit, N. Islam, S. Mondal, and T. Paul, "An integrated social network marketing metric for business-to-business SMEs," *Journal of Business Research*, vol. 150, pp. 73–88, Nov. 2022, doi: 10.1016/j.jbusres.2022.06.006.
  - [12] M. N. A. Raja and S. K. Shukla, "Multivariate adaptive regression splines model for reinforced soil foundations," *Geosynthetics International*, vol. 28, no. 4, pp. 368–390, Aug. 2021, doi: 10.1680/jgein.20.00049.
  - [13] S. Sekhar Roy, R. Roy, and V. E. Balas, "Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid model of MARS and ELM," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 4256–4268, Feb. 2018, doi: 10.1016/j.rser.2017.05.249.
  - [14] M. Graczyk-Kucharska, A. Özmen, M. Szafranski, G. W. Weber, M. Goliński, and M. Sychała, "Knowledge accelerator by transversal competences and multivariate adaptive regression splines," *Central European Journal of Operations Research*, vol. 28, no. 2, pp. 645–669, Jul. 2020, doi: 10.1007/s10100-019-00636-x.
  - [15] A. Özmen, Y. Zinchenko, and G. W. Weber, "Robust multivariate adaptive regression splines under cross-polytope uncertainty: an application in a natural gas market," *Annals of Operations Research*, vol. 324, no. 1–2, pp. 1337–1367, Oct. 2022, doi: 10.1007/s10479-022-04993-w.
  - [16] J. Görtler *et al.*, "Neo: Generalizing Confusion Matrix Visualization to Hierarchical and Multi-Output Labels," Apr. 2022. doi: 10.1145/3491102.3501823.
  - [17] C. Vinante, P. Sacco, G. Orzes, and Y. Borgianni, "Circular economy metrics: Literature review and company-level classification framework," *Journal of Cleaner Production*, vol. 288, p. 125090, Mar. 2021, doi: 10.1016/j.jclepro.2020.125090.
  - [18] I. Tregub, N. Drobysheva, and A. Tregub, "Digital economy: Model for optimizing the industry profit of the cross-platform mobile applications market," in *Advances in Intelligent Systems and Computing*, vol. 850, Springer International Publishing, 2019, pp. 21–28. doi: 10.1007/978-3-030-02351-5\_3.
  - [19] R. J. Vanderbei, *Linear Programming*, vol. 285. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-39415-8.
  - [20] O. Salima, A. Taleb-Ahmed, and B. Mohamed, "An improved Spatial FCM algorithm Based on Artificial Bees Colony," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 1, no. 3, Oct. 2012, doi: 10.11591/ij-ai.v1i3.765.
  - [21] B. Kalayci, A. Özmen, and G. W. Weber, "Mutual relevance of investor sentiment and finance by modeling coupled stochastic systems with MARS," *Annals of Operations Research*, vol. 295, no. 1, pp. 183–206, Aug. 2020, doi: 10.1007/s10479-020-03757-8.
  - [22] A. Özmen and G. W. Weber, "RMARS: Robustification of multivariate adaptive regression spline under polyhedral uncertainty," *Journal of Computational and Applied Mathematics*, vol. 259, no. PART B, pp. 914–924, Mar. 2014, doi: 10.1016/j.cam.2013.09.055.
  - [23] A. Özmen, E. Kropat, and G. W. Weber, "Robust optimization in spline regression models for multi-model regulatory networks under polyhedral uncertainty," *Optimization*, vol. 66, no. 12, pp. 2135–2155, Jul. 2017, doi: 10.1080/02331934.2016.1209672.
  - [24] A. Özmen, İ. Batmaz, and G. W. Weber, "Precipitation Modeling by Polyhedral RCMARS and Comparison with MARS and CMARS," *Environmental Modeling and Assessment*, vol. 19, no. 5, pp. 425–435, Mar. 2014, doi: 10.1007/s10666-014-9404-8.
  - [25] R. Syah, M. Elveny, M. K. M. Nasution, and G. W. Weber, "Enhanced knowledge acceleration estimator optimally with MARS to business metrics in merchant ecosystem," in *2020 4th International Conference on Electrical, Telecommunication and Computer Engineering, ELTICOM 2020 - Proceedings*, Sep. 2020, pp. 1–6. doi: 10.1109/ELTICOM50775.2020.9230487.
  - [26] P. Taylan and G. W. Weber, "CG-Lasso Estimator for Multivariate Adaptive Regression Spline," in *Nonlinear Systems and Complexity*, Springer International Publishing, 2019, pp. 121–136. doi: 10.1007/978-3-319-90972-1\_9.
  - [27] B. L. Gorissen, I. Yanikoglu, and D. den Hertog, "A practical guide to robust optimization," *Omega (United Kingdom)*, vol. 53, pp. 124–137, Jun. 2015, doi: 10.1016/j.omega.2014.12.006.
  - [28] N. Gu, H. Wang, J. Zhang, and C. Wu, "Bridging Chance-Constrained and Robust Optimization in an Emission-Aware Economic Dispatch with Energy Storage," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1078–1090, Mar. 2022, doi: 10.1109/TPWRS.2021.3102412.
  - [29] D. Bertsimas, M. Sim, and M. Zhang, "Adaptive distributionally robust optimization," *Management Science*, vol. 65, no. 2, pp. 604–618, Feb. 2019, doi: 10.1287/mnsc.2017.2952.
  - [30] M. S. Atabaki, M. Mohammadi, and B. Naderi, "New robust optimization models for closed-loop supply chain of durable products: Towards a circular economy," *Computers and Industrial Engineering*, vol. 146, p. 106520, Aug. 2020, doi: 10.1016/j.cie.2020.106520.
  - [31] S. Mahdini, M. Rezaie, M. Elveny, N. Ghadimi, and N. Razmjooy, "Optimization of pemfc model parameters using meta-heuristics," *Sustainability (Switzerland)*, vol. 13, no. 22, p. 12771, Nov. 2021, doi: 10.3390/su132212771.
  - [32] R. Syah, M. K. M. Nasution, E. B. Nababan, and S. Efendi, "Sensitivity Of Shortest Distance Search In The Ant Colony Algorithm With Varying Normalized Distance Formulas," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 19, no. 4, pp. 1251–1259, Aug. 2021, doi: 10.12928/TELKOMNIKA.v19i4.18872.
  - [33] H. Al-Khazraji, A. R. Nasser, and S. Khilil, "An intelligent demand forecasting model using a hybrid of metaheuristic optimization and deep learning algorithm for predicting concrete block production," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 2, pp. 649–657, Jun. 2022, doi: 10.11591/ijai.v11.i2.pp649-657.
  - [34] A. Luque, A. Carrasco, A. Martin, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, Jul. 2019, doi: 10.1016/j.patcog.2019.02.023.

## BIOGRAPHIES OF AUTHORS






**Marischa Elveny**    is an S.T.I. (bachelor's) degree in Information Technology at the Universitas Sumatera Utara, and earned his M.Kom. (master) degree in Informatics Engineering, University of Sumatera Utara. She earned his doctorate (Ph.D.) at the University of Sumatera Utara. Currently working as a lecturer in the Faculty of Computer Science and information Technology at the Universitas Sumatera Utara. Research interests are business intelligence, data science and information technology. She can be contacted via email: marischaelveny@usu.ac.id.






**Mahyuddin K. M. Nasution**    is Professor from Universitas Sumatera Utara, Medan Indonesia. Worked as a Lecturer at the Universitas Sumatera Utara, fields: i) Mathematics, computer science, Computer, and Information Technology; and ii) Education: Drs. Mathematics (USU Medan, 1992), MIT Computers and Information Technology (UKM Malaysia, 2003); Ph.D. in Information Science (Malaysian UKM). He can be contacted via email: mahyuddin@usu.ac.id.






**Muhammad Zarlis**    is a Professor at Bina Nusantara University with, Study program of Master of Information System Management, he is the fields of Intelligent Systems and Big Data, Algorithm and Programming and Numerical Method and Optimization. Studied master at the Universitas Indonesia and Ph.D. at the University Sains Malaysia. He can be contacted via email: Muhammad.zarlis@binus.edu.



**Syahril Efendi**    is a doctorate at the Faculty of Computer Science and Information Technology, Universitas Sumatera Utara majoring in Information Science and Operations Research. pursued undergraduate education at the Universitas Sumatera Utara and earned a doctorate at the Universitas Sumatera Utara. The field of science occupied is Research Operation Information Science. Should be contacted via email syahrilefendi@usu.ac.id.



**Rahmad B. Y. Syah**    Received head of Excellent Centre of Innovations and New Science (PUIN) Universitas Medan Area. His research interests are modeling and computing, artificial intelligence, Data Science and Computational Intelligent. He is a member of the IEEE, Institute for System and Technologies of Information, Control, and Communication, IAENG (International Association of Engineers). He should be contacted at email: rahmadsyah@uma.ac.id.