Troubleshooting Customer Behaviour Against Merchants with Adaptive Multivariate Regression

Marischa Elveny*
Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Padang Bulan USU, Medan, Indonesia. E-mail: marischaelveny@usu.ac.id

Mahyuddin KM Nasution
Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Padang Bulan USU, Medan, Indonesia.

Muhammad Zarlis
Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Padang Bulan USU, Medan, Indonesia.

Syahril Efendi
Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Padang Bulan USU, Medan, Indonesia.

Received April 05, 2021; Accepted August 02, 2021
ISSN: 1735-188X
DOI: 10.14704/WEB/V18I2/WEB18333

Abstract

Business intelligence can be said to be techniques and tools as acquisition, transforming raw data into meaningful and useful information for business analysis purposes. This study aims to build business intelligence in optimizing large-scale data based on e-metrics. E-metrics are data created from electronic-based customer behavior. As more and more large data sets become available, the challenge of analyzing data sets will get bigger and bigger. Therefore, business intelligence is currently facing new challenges, but also interesting opportunities, where can describe in real time the needs of the market share. Optimization is done using adaptive multivariate regression that can be address high-dimensional data and produce accurate predictions of response variables and produce continuous models in knots based on the smallest GCV value, where large and diverse data are simplified and then modeled based on the level of behavior similarity, basic measurements of distances, attributes, times, places, and transactions between social actors. Customer purchases will represent each preferred behaviour and a formula can be used to calculate the score for each customer using 7 input variables. Adaptive multivariate regression looks for customer behaviour so as to get the results of cutting the deviation which is the determining factor for performance on the data. The results show there are strategies and information needed for a sustainable business. Wheremerchants who sell fast food or food stalls are more in demand by customers.

Keywords

Business Intelligence, E-Metrics, Behaviour, Merchant, Adaptive Multivariate Regression.

Introduction

The main objective of business intelligence (BI) is to extract strategic knowledge from information provided by various data sources to provide assistance in the decision-making process and achieve the company's strategic goals. In recent years, BI-oriented data processing and analysis has developed (Bordeleau, F. E., Mosconi, E., & de Santa-Eulalia, L. A, 2020).

A newer approach that fits the current needs of big data processing, focuses more on the speed and immediacy of information, data processing in streaming and, depending on the model designed, incorporates batch analysis processes to derive a knowledge model. In this way, they can offer new analytical data and enriched information from the model (Bentley, D, 2017).

BI can be used to support various business decisions ranging from operational to strategic (Rasmussen, N.H., Goldy, P.S., & Solli, P.O, 2002). BI plays an important role in improving organizational performance by identifying new opportunities, highlighting potential threats, revealing new business insights, and improving the decision-making process. One of Indonesia's strengths is micro, small and medium enterprises (MSMEs), which dominate the Indonesian economy (Turban, E, 2011) (Zheng, Jack, 2019).

In competition and success, information requires strategies and plans, and in these plans and plans, competitive organizations where electronic indicators play a role will not be able to ignore these strategies and plans (Melo, P. N., & Machado, C, 2019). Electronic metrics or also known as electronic metrics are data created based on electronic consumer behaviou (Dijkman, R., Dumas, M., Van Dongen, B., Käärik, R., & Mendling, J, 2011).

Retrieval of information in competitive organizations implies similarity, and this similarity is used to find relationships between competitive organizational behaviour. The similarity of two objects is measured using the concept of closeness, which is used to obtain a finite value. There are many similarities in size, but entering the sequence of object components requires a special approach (M Elveny., R Syah., M.K.M, Nasution, 2020) (M.K.M. Nasution, O.S. Sitompul, S. Nasution, H. Ambarita, 2017). This method is carried out by analysing customer data based on similarities in behaviour, basic measures of distance, attributes, time, location, and transactions. The relationship between stakeholders can be used to trace information sources, and can also be used to predict information sources for social behaviour.

In terms of behavioral data metrics (Kunze, Matthias, Matthias Weidlich and Mathias Weske, 2011) introduced metrics which are very suitable for measuring process trends using behavior profiles. The indicators used were evaluated successfully using the assumptions used to assess human trends. Meanwhile, in terms of method optimization, (Özmen, A., Weber, G. W., & Batmaz, İ, 2010) A strong CMARS system resolves conflicts by optimizing the sensitivity of problem parameters to irregularities. Data uncertainty creates unclear constraints. To solve this problem, the best amplification technique is used to solve the uncertainty in the data. Optimization is a process to get the desired result or design and create the best.

Based on previous research, the purpose of this study is to solve problems in customer behaviour towards competitive merchants using adaptive multivariate regression to produce an optimal model that can use electronic measurement data and this data was not carried out in previous studies.

Formulation of the Problem

Information requires strategies, methods and tools that competitive organizations cannot ignore through electronic measurement data. On the one hand, not all customer profiles are alike, but every transaction may have an anomaly. On the other hand, the methods used can be adapted to the different behaviours carried out by competing organizations. To optimize customer data, reliable nom-parametric regression is used to make predictions from electronic measurement data.

Material and Method

There are several stages that are carried out. In data processing using multivariate adaptive regression spline (MARS) (Özmen, A., Batmaz, İ., & Weber, G. W, 2014) (Özmen, A., & Weber, G. W, 2014). MARS functions as data validation and looks for customer behavior in order to get the result of cutting the deviation which is the determining factor (Jaklič, J., Grublješič, T., & Popovič, A, 2018). The model obtained is tested by grouping it based on similar merchants and merchants of other types. The last stage is the visualization of the data obtained based on this grouping which produces a graph that can show customer behavior towards merchants which will later be used to predict sustainable business.

Business Intelligence

In general, business intelligence has several processes that will be carried out, namely data sets which are collections of various raw data. Purify information, data that has been shared and selected based on their respective functions. Data storage, filtered data is stored in a specific place and ready to be analyzed (Rikhardsson, P., & Yigitbasioglu, O, 2018) (Torres, R., Sidorova, A., & Jones, M. C, 2018). Data analysis, this process involves various ways of analysis to extract knowledge and information. Data visualization, results are presented and presented in different ways that can be understood as decision support (Syah, R., Nasution, M. K. M., Nababan, E. B., & Efendi, Syaril, 2020).

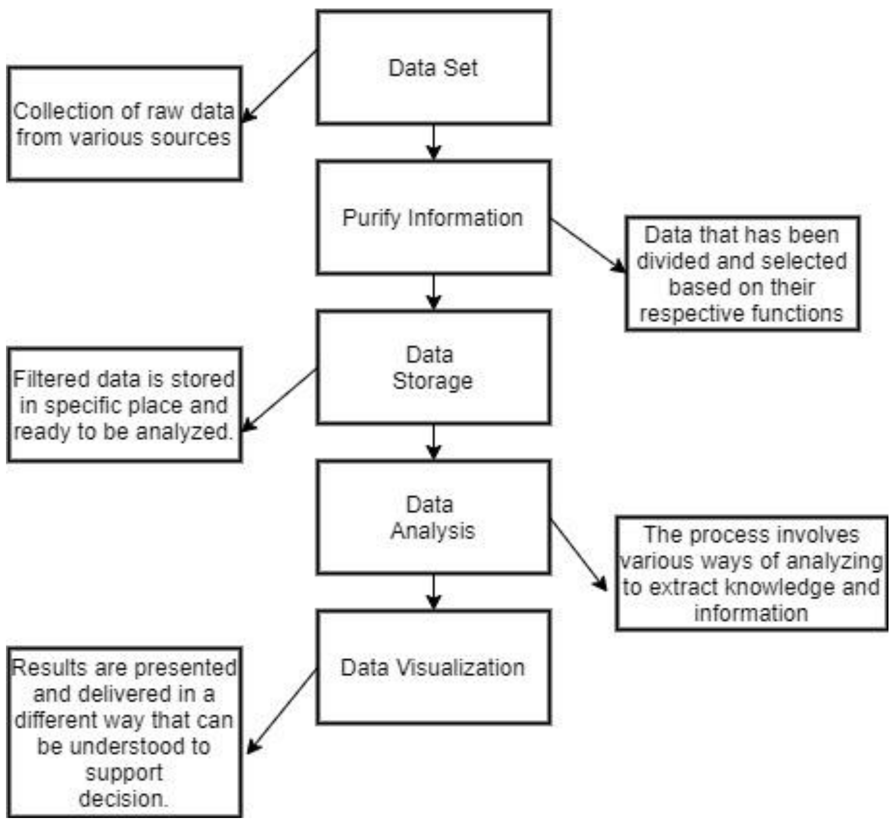

Figure 1 General Process Business Intelligence

E-Metrics

Variants of activity and total merchants vary depending on the type of business actor, especially digital business (R. Syah, M. Elveny, M. K. M. Nasution and G. W. Weber, 2020). In electronic-based activities, it can be seen that the user's transaction habits make it easier to see the condition of the opinion. With many sellers using digital business development electronically, you can track the term revenue or what is called profit margin more easily. Likewise, because of the large number and variety of changes, business uncertainty is increasingly difficult to predict (Syah, R., Nasution, M. K., Nababan, E. B., & Efendi, S, 2020). At this point, future commercial options will be grouped and classified (Özmen, A, 2010). You will get outliers or differences, which can be used as a profit. Figure 2 shows customer activity when transacting based on e-metric data.

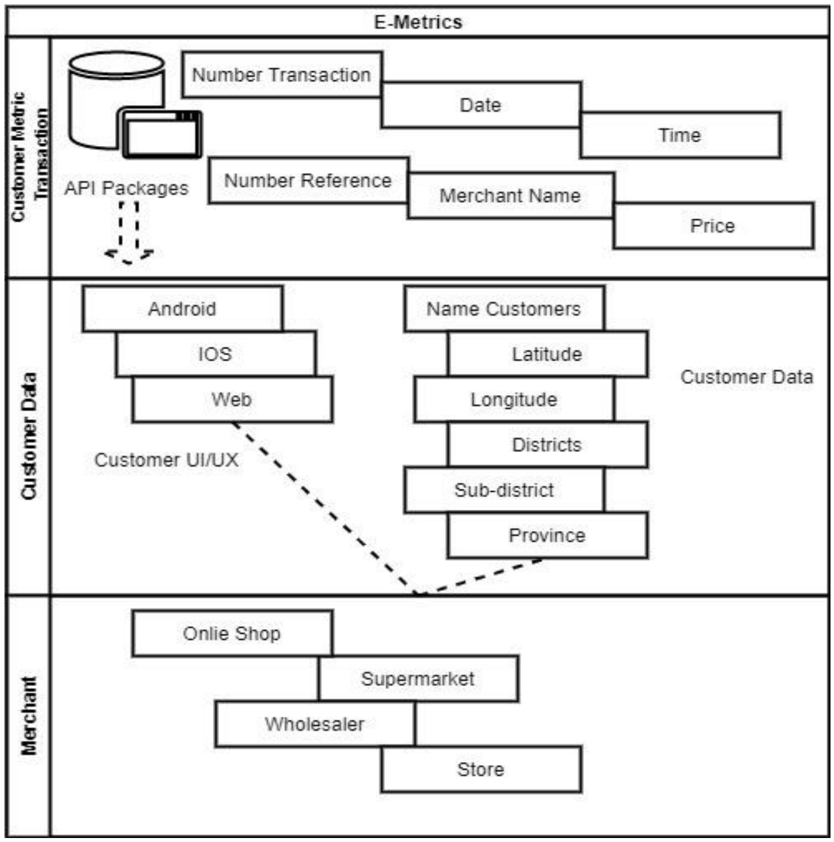

Figure 2 E-Metrics

MARS (Multivariate Adaptive Regression Spline)

MARS is an adaptive mechanism, because the selection of database-based functions is specific to the problem at hand (Weber, G.W., Batmaz, İ., Köksal, G., Taylan, P., & Yerlikaya-Özkurt, F, 2012). The specific advantage of MARS lies in its ability to estimate the basis of the function as a result of which additional and interactive effects according to predictors are allowed to predict response variables. In this case MARS uses expansion in piecewise linear form (Weber, G. W., Tezel, A., Taylan, P., Soyler, A., & Çetin, M, 2008).

$$[+(x - T)]_+, [-(x - T)]_+, \qquad (3.1)$$

Where [q] +: = max {0, g}
τ = univariate node.

Each function is straight linear, with the vertex at the value of r, and the corresponding function pair is called the reflected pair.

The general model of the relationship between the predictor variable and the response objective is to construct the reflected pairs for each predictor xj (j = 1, 2 ..., p) with p-dimensional knots $\tau_i = (\tau_{i,1}, \tau_{i,2}, ... \tau_{i,p})$ T at xi = (xi, 1, xi, 2, ... xi, p) T or just adjacent to each data vector ẋi = (ẋi, 1, ẋi, 2, ... ẋi, p) T (i = 1, 2, ..., N) of the predictors (Weber, G. W., Çavuşoğlu, Z., & Özmen, A, 2012).

It can be assumed $\tau_{i,\,j} \neq \dot{x}_i$, j for all i and j, to prevent indifference in the optimization problem of this research later. Actually, we can select vertices $\tau_{i,\,j}$ further than the predictor value $\dot{x}_i$ j, if there is a position that promises better data fitting (Weber, G.-W., Batmaz, I., Köksal, G., Taylan, P., and Yerlikaya, F, 2009).

$$X_j = X + \mu_j (j = 1,2, \dots, m) \quad (3.2)$$

Result and Discussion

Variable Input

There are 3 types of data used, namely registered customers, active customers and passive customers. Registered customers are customers who have never made a transaction at all. Active customers are customers who frequently make transactions, while passive customers are customers who have made transactions but not frequently. In testing the data using MARS, interaction and basis functions are carried out. Where the interaction is the result of cross multiplication between correlating variables. And the basis of the function is a function used to explain the relationship between the response variable and the predictor variable.

Table 1 Variable Input

Predictor (X)	Variable Input
X1	No Transaction
X2	Date
X3	Time
X4	No Reference
X5	Customer's Name
X6	Transaction
X7	Merchant Name

These variables function as markers and links to find relationships between related data. Data were validated against 7 inputs, where each predictor variable x must be determined based on the independent variable and the dependent variable. Then we will use MARS according to the number entered for search based on information optimization. Test each predictor variable (X1), (X2), (X3), (X4), (X5), (X6), (X7) using the best value and the maximum iteration value. Use MARS to optimize information with a computer. Check each predictor variable (X1), (X2), (X3), (X4), (X5), (X6), (X7) using the best value and the maximum value from the iteration calculation results.

Training Data

Furthermore, training will be conducted for the sub set merchant name and the sub set for purchases. The sub set is intended to find the names of registered customers and make transactions.

With respect to using any amount claimed f_α. This study notes that α talk about some of the complexity of estimating. To estimate the optimal value α, generalized cross-validation (GCV) can be used (Syah, R., Nasution, M.K., Elveny, M., & Arbie, H, 2020).

$$\mathrm{GCV} := \frac{1}{N}\frac{\sum_{i=1}^{N}\left(y_i - \hat{f}_\alpha(x_i)\right)^2}{\left(1-\tilde{C}(\alpha)/N\right)^2} \qquad (4.1)$$

Table 2 Value of Data Validation Using GCV

N Inputs	Effective Parameters	GCV
3	20,20	139,180,185,282
2	17.80	140,018,911,541
2	15.40	140,290,949,538
1	13.00	140,952,126,568
1	10.60	149,949,236,267
1	8.20	182,817,946,513
1	5.80	189,919,290,425
1	3.40	202,187,820,719
0	1.00	218,412,528,000

Table 3 Table of Customer Behavior Subset Based on Variable Indicators

Indicator	Type	Level	Score
Type of Transaction	Random	50	1500; 3000; 3001; 4200; 5500; 6500; 7000; 8000; 9000; 9500; 10000; 12000; 13000; 14500; 15000; 77000; 71000; 16000; 18500; 12000; 19000; 20000; 24000; 25000; 27500; 29500; 30000; 31000; 56000; 36000; 40000; 42500; 45000; 48000; 50000; 60000; 70000; 10000; 72000; 75000; 85000; 90000; 100000; 140000; 185000; 15000; 17000; 78000; 100000; 110000
Merchant Name	Fixed	29	Grilled ash, various fried foods, various pharmacies, Bandung chicken porridge, bu neng burger, indomaret, asia mega mas, kfc adam malik, bu neng, kede john, gallon rendy, kfc mareland, kfc center point, ir one, kede yusuf, noodles aceh andrge, KFC Adam Malik Medan; KFC Asia Mega Mas Medan; KFC Btc Mareland, KFC Cemara Asri Medan, KFC Simpang Mataram Medan, Aminah, Fried Sausages alpresto, Toko Dedi, Warung Abas, Warung Simpang Tiga.

Function Base

The Basis Function is a function used to explain the relationship between the response variable and the predictor variable.

The function base can be searched by (R. Syah, M. Elveny and M. K. M. Nasution, 2020):

$$\wp : \left\{ (x_j - T)+, (T - x_j) + \mid T \in \{x_{1,j}, x_{2,j}, \dots, x_{N,j}\},\ J \in \{1,2,\dots,p\} \right\}, \quad (4.2)$$

Thus, we can represent f (x) with a linear combination constructed by the set p and by the intercept respectively θo, so that it takes the following form:

$$y = \theta_0 + \sum_{m=1}^{M} \theta_m \psi_m(x) + \epsilon, \quad (4.3)$$

Here, ψ_m (m = 1, 2, ..., M) is the BF of p
ψ is the unknown coefficient for the m-base function (m = 1,2, ..., M)

One set of vertices that meet the conditions $\tau_{i,j}$ are assigned separately for each dimension of the predictor variable and are selected to coincide with the level of the predictor represented in the data. The interaction BF is created by multiplying the existing BF by a truncated linear function involving the new variable. In this case, the existing BF and the newly created BF interactions are used in the MARS approach. Provided that the observations are represented by the data $\boldsymbol{x}_i$, $\boldsymbol{y}_i$ (i = 1, 2,..., N), the BF to m form can be written as follows (Özmen, A., Weber, G. W., Batmaz, İ., & Kropat, E, 2011):

$$\psi_m(x) := \prod_{j=1}^{K_m} \left[S_{k_j^m} \cdot \left(x_{k_j^m} - T_{k_j^m} \right) \right] +, \quad (4.4)$$

Where:
Km: the sum of truncated linear functions multiplied in BF to-*m*,
$x_{k_j^m}$: the predictor variable corresponding to the *j* intersects the linear function in BF to-*m*,
$T_{k_j^m}$: the value of the node according to the variable

In this case, 11 experiments will be carried out on a function basis which will later get the results of the upper, middle and lower thresholds. The function basis to be carried out is 1,2,3,5,7,9,11 to solve the optimization problem. This research can select the node value as the data point, but because the value is very close to each of the various variables, then to find competitive predictions of the need for an optimum base function model.

Table 4 Deviation Cutting Performance Results

Percentile	Joint N	Mean	MSE
90%	640	12242	1433
80%	569	12519	638
70.00%	498	11511	169
60%	427	10493	105
50% Median	356	11291	71
40%	285	12006	40
30%	214	12006	40
20%	143	12006	40
10%	72	12006	40
5%	36	12006	40
1%	8	12006	40
99.86%	-1	710	9640
99.30%	-5	706	7138
98.58%	-10	701	5766
96.49%	-25	868	4163
92.97%	-50	661	3453
85.94%	-100	611	2268

Grouping

Based on the resulting model, it shows the competitive relationship of each merchant that has been generated from the equations X1, X2, X3, X4, X5, X6, X7. Observations are made based on the number of customer behaviour activities by looking at the intensity of transactions against merchants. The grouping is carried out based on similar merchants and merchants of different types. Figure 3 shows customer behaviour towards a merchant based on input variables. Where each behaviour is different.

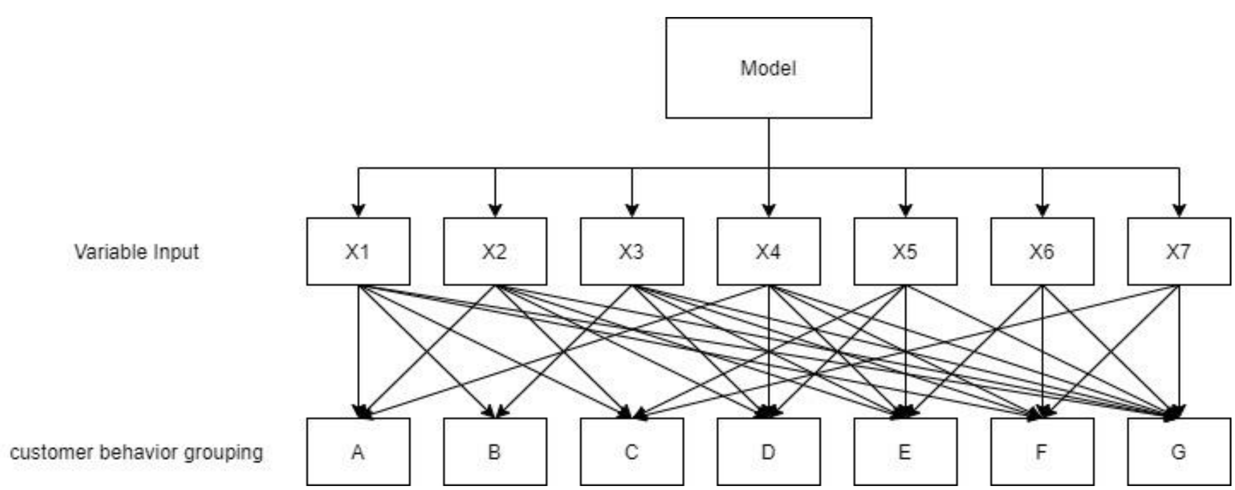

Figure 3 Customers behavior grouping

Result

From the results of behavior based on predictors, a match is sought by looking at obs, transaction value, similarity, resid and std resid. It is intended to generate predictions based on transactions made by customers against merchants. The result is that if the value is F then it is not suitable, while the value is T then it is appropriate.

Table 5 Predictor Matching Results

Obs	Transaction Value	Similarity	Resid	Std Resid		
25	1,50000	1,50000	0.000006	0.000133		F
100	8,50000	2,95710	5,54290	2,863172	T	
118	1.00000	3,27713	6.72287	3.822056	T	
122	7,50000	2,95710	4,54290	2.346624	T	
126	7.00000	2,95710	4,04290	2,088350	T	
165	3000,000	2999,999	0.000006	0.000108		F
168	1,40000	2,71694	1,12831	5.851783	T	
189	7,50000	2.79687	4,70313	2.456029	T	
239	6.00000	1,97500	4,02500	2.111985	T	
265	6.00000	1.9750	4,02500	2.111985	T	
273	5500,000	5499,999	0.000006	0.000108		F
322	4200,000	4199,999	0.000006	0.000107		F
329	1.85000	2,59032	1,59097	8,209327	T	
393	6500,000	6499,999	0.000006	0.000142		F
402	1,30000	1,30000	0.000006	0.000142		F
421	1,50000	1,50000	0.000006	0.000142		F
447	9.00000	2,37340	6,62660	3,408107	T	
480	1,50000	1,50000	0.000006	0.000142		F

Table 6 Maximum and Minimum Value

Merchant Name	Value
Abu Bakar	0.525
Aneka Gorengan	0.584
Apotik Aneka	0.921
Bubur Ayam bandung	0.473
Burger Bu neng	0.718
Indomaret	0.682
Asia Mega Mas	0.973
Kede John	0.934
Gallon rendy	0.031
KFC	0.573
IR One	0.673
Kede John	0.404
KFC Adam Malik	0.412
KFC Asia Mega Mas Medan	0.437
KFC Btc Mareland	0.457
KFC Cemara Asri Medan	0.736
KFC Center Point	0.469
KFC Simpang Mataram Medan	0.372
Aminah	0.937
Sosis Goreng Alpresto	0.921
Toko Dedi	0.894
Warung Abas	0.021
Warung Simpang Tiga	0.981

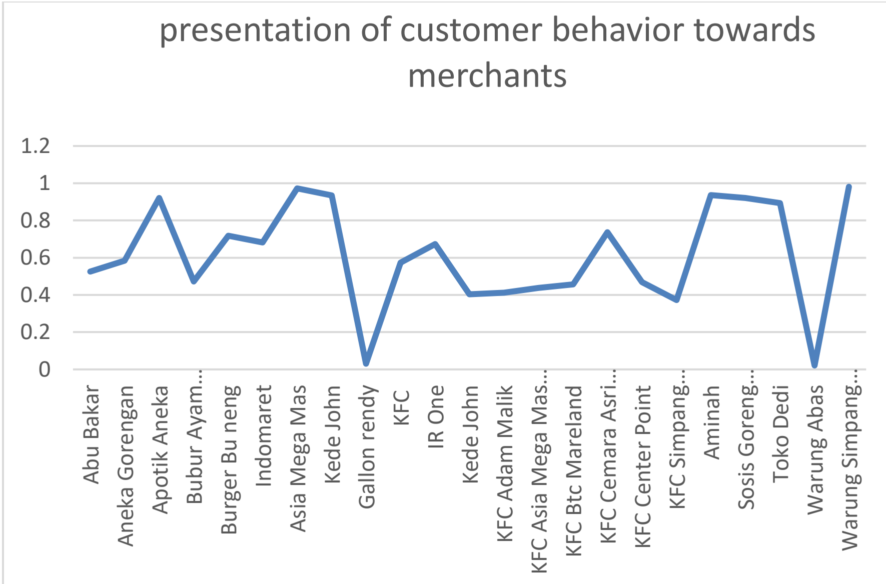

Figure 4 Graph of Customer Behavior Results towards Merchants

Based on Figure 4, the results show that merchants who sell fast food or food stalls are more in demand by customers.

Conclusion and Future Work

The results of this study can be obtained by various studies of information poured into the knowledge base, namely in the form of data used in the form of e-metric data as the basis for searching customer behavior towards competitive merchants, then the data is tested using MARS as a form to solve dimensional data problems. high. Through MARS the data is generated to find the maximum value with upper, middle and lower thresholds using the basis of functions, as well as interactions that function as a match between input variables so that there is a simplification of behavior. The results achieved are in the form of maximum value information based on behavior towards competitive merchants to be used as decision making in sustainable business.

References

Bordeleau, F.E., Mosconi, E., & De Santa-Eulalia, L.A. (2020). Business intelligence and analytics value creation in Industry 4.0: a multiple case study in manufacturing medium enterprises. *Production Planning & Control*, *31*(2-3), 173-185.

Bentley, D. (2017). Business Intelligence and Analytics. *Internet, link: https://www. pdfdrive. com/business-intelligence-and-analytics-e56416503.html*

Rasmussen, N.H., Goldy, P.S., & Solli, P.O. (2002). *Financial business intelligence: trends, technology, software selection, and implementation*. John Wiley & Sons.

Turban, E. (2011). *Business intelligence and analytics: Systems for decision support, global edition.* Pearson.

Zheng, Jack. (2019). *Business Intelligence and Analytics a Comprehensive Overview*.

Melo, P.N., & Machado, C. (Eds.). (2019). *Business Intelligence and Analytics in Small and Medium Enterprises.* CRC Press.

Dijkman, R., Dumas, M., Van Dongen, B., Käärik, R., & Mendling, J. (2011). Similarity of business process models: Metrics and evaluation. *Information Systems, 36*(2), 498-516.

Elveny, M., Syah, R., & MKM, N. (2020). Taxonomy business model to customer behaviour in e-metric ecosystem. *Journal of Advanced Research in Dynamical and Control Systems, 12*(7), 99-104. https://doi.org/10.5373/JARDCS/V12I7/20201989

Nasution, M.K., Sitompul, O.S., Nasution, S., & Ambarita, H. (2017). New similarity. In *IOP Conference Series: Materials Science and Engineering, 180*(1).

Kunze, M., Weidlich, M., & Weske, M. (2011). Behavioral similarity–a proper metric. *In International Conference on Business Process Management*. Springer, Berlin, Heidelberg, 166-181.

Özmen, A., Weber, G.W., & Batmaz, İ. (2010). The new robust CMARS (RCMARS) method. *Vectors, 1*, 362-368.

Özmen, A., Batmaz, İ., & Weber, G. W. (2014). Precipitation modeling by polyhedral RCMARS and comparison with MARS and CMARS. *Environmental Modeling & Assessment, 19*(5), 425-435.

Özmen, A., & Weber, G.W. (2014). RMARS: robustification of multivariate adaptive regression spline under polyhedral uncertainty. *Journal of Computational and Applied Mathematics, 259*, 914-924.

Jaklič, J., Grublješič, T., & Popovič, A. (2018). The role of compatibility in predicting business intelligence and analytics use intentions. *International Journal of Information Management, 43*, 305-318.

Rikhardsson, P., & Yigitbasioglu, O. (2018). Business intelligence & analytics in management accounting research: Status and future focus. *International Journal of Accounting Information Systems, 29*, 37-58.

Torres, R., Sidorova, A., & Jones, M.C. (2018). Enabling firm performance through business intelligence and analytics: A dynamic capabilities perspective. *Information & Management, 55*(7), 822-839.

Syah, R., Nasution, M.K.M., Nababan, E.B., & Efendi, S. (2020). Knowledge Acceleration Estimator (KAE) Model to Customer Behavior Using Business Metrics. *Journal of Theoretical and Applied Information Technology, 98*(8), 1301-1311.

Syah, R., Elveny, M., Nasution, M.K.M., & Weber, G.W. (2020). Enhanced Knowledge Acceleration Estimator Optimally with MARS to Business Metrics in Merchant Ecosystem. *In 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)*, 1-6. https://doi.org/10.1109/ELTICOM50775.2020.9230487.

Syah, R., Nasution, M.K., Nababan, E.B., & Efendi, S. (2020). Optimization Model Under Uncertainty for Competitive Merchant to Customer Behavior in Financial Technology

Using Business Metrics. *Journal of Computational and Theoretical Nanoscience, 17*(7), 3264-3270.

Özmen, A. (2010). Robust conic quadratic programming applied to quality improvement-A robustification of CMARS (Master's thesis).

Weber, G.W., Batmaz, İ., Köksal, G., Taylan, P., & Yerlikaya-Özkurt, F. (2012). CMARS: a new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimization. *Inverse Problems in Science and Engineering*, *20*(3), 371-400.

Weber, G.W., Tezel, A., Taylan, P., Soyler, A., & Çetin, M. (2008). Mathematical contributions to dynamics and optimization of gene-environment networks. *Optimization*, *57*(2), 353-377.

Weber, G.W., Çavuşoğlu, Z., & Özmen, A. (2012). Predicting default probabilities in emerging markets by new conic generalized partial linear models and their optimization. *Optimization*, *61*(4), 443-457.

Weber, G.W., Batmaz, İ., Köksal, G., Taylan, P., & Yerlikaya-Özkurt, F. (2012). CMARS: a new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimization. *Inverse Problems in Science and Engineering*, *20*(3), 371-400.

Syah, R., Nasution, M.K., Elveny, M., & Arbie, H. (2020). Optimization Model for Customer Behavior with MARS and KYC System. *Journal of Theoretical and Applied Information Technology, 98*(13), 2616-2627.

Syah, R., Elveny, M., & Nasution, M.K. (2020). Performance knowledge acceleration optimization with MARS to customer behavior in merchant ecosystem. *In 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT)*, 178-182. https://doi.org/10.1109/MECnIT48290.2020.9166604

Özmen, A., Weber, G.W., Batmaz, İ., & Kropat, E. (2011). RCMARS: Robustification of CMARS with different scenarios under polyhedral uncertainty set. *Communications in Nonlinear Science and Numerical Simulation, 16*(12), 4780-4787.

Kravets, P., Burov, Y., Lytvyn, V., & Vysotska, V. (2019). Gaming method of ontology clusterization. *Webology, 16*(1), 55-76.