Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 5 (2023) 1033 - 1039 ISSN Media Electronic: 2580-0760

Precision Marketing Model using Decision Tree on SME e-commerce Case Study Orebae.com

Fadil Indra Sanjaya[1], Anna Dina Kalifia[2]
[1, 2] Informatika, Fakultas Sains & Teknologi, Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia
[1]fadil.indra@staff.uty.ac.id, [2]anna.dina.kalifia@staff.uty.ac.id

Abstract

The development of the industrial world towards industry 4.0 has resulted in changes in the lifestyle of the wider community in carrying out their activities through digital media, one of which is shopping. This has an impact on the emergence of many business actors in the e-commerce field which brings its own challenges to stay alive and face the competition. The demands for innovation in competitive competition are also increasingly diverse with various approaches ranging from technology, social science, management science and even artificial intelligence. One form of innovation that is widely carried out by e-commerce today is looking for an ideal and effective form of marketing, where the form of marketing itself is considered less able to accommodate e-commerce needs. One form of real innovation in finding the ideal and effective marketing is with precision marketing. Precision marketing itself is marketing that is carried out by utilizing data where consumers are the center of preference for data collection. In fact, many e-commerce companies that were launched were unable to keep up with the competition because they were unable to develop marketing strategies and eventually went bankrupt. Therefore, we need a special way to bridge these problems so that e-commerce can stay alive, especially for e-commerce classified as Small and Medium Enterprises (SMEs). This research will focus on developing a precision marketing model in SME e-commerce, namely orebae.com which can be used as a tool in developing marketing strategies. This research was conducted using a machine learning approach by adopting a decision tree algorithm. From the results of this study, showed that precision marketing model for orebae.com according to customer preferences, can be used to increase the number of orebae.com sales and to reduce marketing cost.

Keywords: precision marketing; SME; machine learning; customer preference; supervised learning; decision tree

1. Introduction

The development of the industrial world towards industry 4.0 has a significant impact on digital transformation, especially for entrepreneurs [1]. As culture changes into digital culture in the wider community, resulting in changes in people's lifestyles that do more activities via digital media. The routine activity that is often done is shopping, which is currently accommodated with online media, one of which is through e-commerce. This brings its own challenges for business actors in the e-commerce field in conducting competition. Often with the emergence of much e-commerce, more and more innovations are successfully generated to win the competition. One form of innovation that is often carried out by e-commerce is to find an ideal and effective form of marketing, where the traditional form of marketing itself incapable of accommodating the needs of e-commerce so as not to be eliminated from the competition and in maximizing profits [2].

The fact is that many SME e-commerce are not able to keep up with the competition where consumers prefer to shop at large e-commerce, which ultimately results in the elimination of SME e-commerce from the competition. The biggest mistake of SME e-commerce is the lack of attention to the role of data as a key factor that can be used in formulating strategies, one of which is marketing strategy. Marketing activities themselves cannot be separated from data [3]-[9]. A large amount of marketing data contains a lot of valuable customer information [10], [11]. Precision marketing itself is marketing that is carried out by utilizing data where consumers are the center of their preferences [12]. Through the collection, processing and analysis of this data, precision marketing can be realized to target consumers according to their needs and which is able to have an impact on reducing marketing costs and increasing marketing efficiency [13]. To extract valuable information from a set of data, data mining technology and machine learning have been widely

Accepted: 23-09-2022 | Received in revised: 27-03-2023 | Published: 30-09-2023

applied with several methods such as decision trees, association rules, artificial neural networks, and other methods [14]. Therefore, in this study, researchers will propose an alternative approach by applying machine learning to supervised learning with the Decision tree algorithm to develop a precise marketing model for SME e-commerce orebae.com. This approach is considered by researchers to be better in formulating marketing strategies because it is sourced from consumer preference data which is expected to be able to create marketing strategies that are close to consumer needs. This research hopefully will be able to become a solution for SME e-commerce, especially orebae.com and other SME e-commerce.

In recent years, precision marketing issues have received much attention and researchers have begun to conduct research in this field with various approaches using various techniques and methods. Several similar studies in the realm of precision marketing [3] include using the ADIMA (Attention Interest Desire Memory Action) Model with a combination of artificial neural networks to perform model analysis and evaluation and K-mean Clustering for model optimization based on customer data stored in the database. data. Another study by [4], [15] with the RFM (Recency, Frequency, Monetary, Value) model where customer data is collected based on the identification of their shopping habits such as shopping frequency, amount of money spent etc. then used for grouping customers according to their characteristics. Another study in research [10] with the application of AISAS (Attention, Interest, Search, Action, and Share) model with the AHP method on consumer purchase history data. Other research was also conducted by [16] using the User Interest Graph (UIG) model on user interest and there is also research by [17], [18] based on big data theory with several types of data sources including customer databases, social media, online platforms, and market analysis. From the description, a comparison of studies in the case of precision marketing is shown in the Table 1.

Table 1. Related Work

Model	Data	Researcher
RFM Model, K-Mean Clustering	Sales transaction history data in database	Ina Maryani dan Dwiza Riana (2017)
Big Data Model	Social media, online platforms, and market-analysis	Yushui Xiao, dan Feng Ling (2019)
ADIMA Model, K-Mean Clustering	Customer shopping history data in database	Jianfeng Cheng (2020)
User Interest Graph Model	online records (number of posts, number of likes, number of comments)	Zhiguo Zhu, dkk. (2020)
AISAS Model, AHP	Customer shopping history data in database	Wu Jun, dkk. (2021)
RFM Model, K-Mean Clustering, Decision Tree	Sales transaction history data in database	Sularso Budilaksono, dkk. (2021)

Looking at comparisons of the research on Table 1, it is very possible for this research to be carried out. In this study, researchers will take a different approach in developing a precision marketing model, namely by using decision tree algorithm in classifying customer habits on customer historical data including the attributes of age, income, occupation, and educational background. This research also analyzes the effects of precision marketing with the application of the resulting model. In the result, the model will be used in supporting marketing program within one month.

2. Research Methods

This research will be carried out in several stages, which have been designed by researchers, which can be seen in Figure 1.

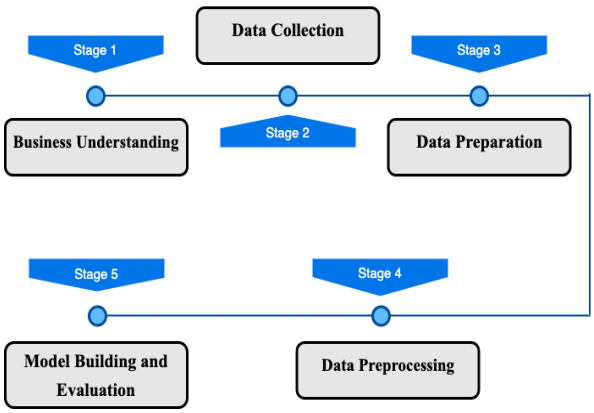

Figure 1. Research Stages

Figure 1 shows that research stage will begin with business understanding, data collection, data preparation, data preprocessing, model building and evaluation.

2.1 Business Understanding

At this stage the researcher conducts some preliminary research discussions to understand the business. This stage consists of a very precise problem specification along with the method of evaluating the achievement of the objectives shown in Table 2.

DOI: https://doi.org/10.29207/resti.v7i5.4531

Table 2. Specification of problem

Problem	Key Success	Factors
Sales cannot meet target	New Marketing Approach	Customer needs
Must produce all products in quantities that do not meet market needs	Product demand forecasting	Market needs
Decision making is not as expected	Decision maker using old method	Innovation to find new method

2.2 Data Collection

The research data used is customer historical data such as age, income, occupation, and educational background. The data was obtained by orebae.com from regular surveys on their customers. The results of the survey will be used usually for the owner and management to see opportunities and in return they will provide prizes or promos for those who are willing to fill out the survey. In the fact, the data that was obtained cannot be handled or processed properly resulting in inaccurate decisions. The survey data that will be used are data from 2020 to 2022. The size of the data that successfully collected was 5832, but for this study we only used total 5000 data. As for sample data will be shown in Table 3.

Table 3. Sample Data

Id	Age	Income	Edu	Occup	label
8	50	3	5	3	0
9	35	2	6	2	0
10	34	2	3	3	1

For the categorization of the attributes of which can be explained as follows:

Occupation
- 1 = jobless
- 2 = state officer
- 3 = private
- 4 = teacher
- 5 = agriculture
- 6 = religion
- 7 = enterpreneur

Education
- 1 = high school level and below
- 2 = diploma (D3) and bachelor (S1)
- 3 = above bachelor (S2, S3)

Income
- 1 = less equal to 1.5 million
- 2 = greater than 1.5 million and less than 5 million
- 3 = less equal to 5 million

Potential Customer
- 0 = buy once or less than once
- 1 = buy more than once

2.3. Data Preparation

At this stage, data cleaning is carried out, from duplicate data, missing values, changing data types to numeric and removing unused features. From the findings, there is no missing data or duplicates data, and data already in numerical form, mean while for the features that are not used and must be deleted are Id.

2.3 Data Preprocessing

The preprocessing stage is carried out by performing Exploratory data analysis (EDA), Univariate Analysis and bivariate-multivariate analysis.

2.3.1 Exploratory Data Analysis (EDA)

Exploratory Data Analysis refers to the critical process of conducting preliminary investigations on data to find patterns, find anomalies, test hypotheses, and check assumptions with the help of summary statistics and graphical representations. The results of the EDA conducted found that: Customers are in the age range of 23 – 67; Most of the customers come from the income group of 5 million and above; Most customers come from high school education and below; Most customers come from entrepreneurs; and Customers have bought more than once on oreBae only 480 people.

2.3.2 Univariate Analysis

Univariate analysis is the simplest form of data analysis. The main purpose of this analysis is almost the same as EDA, which is to retrieve data, summarize the data, and find patterns in the data. At this stage, check the distribution of the data by grouping the data shown in Figure 2.

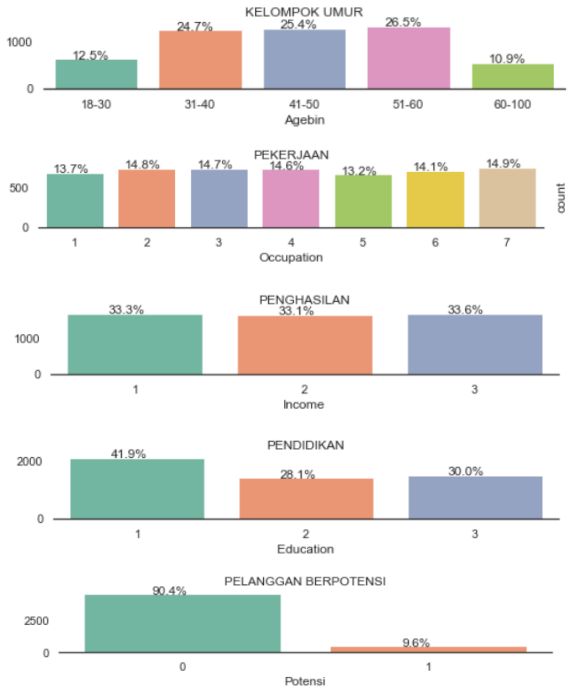

Figure 2. Data Group

DOI: https://doi.org/10.29207/resti.v7i5.4531

Based on the analysis, it was found that the most orebae.com customers were aged 51-60, 41-50 and 31-40. And for customers who have the potential to become regular customers are those with income of more than 1.5 milllion and less than 5 million

2.3.3 Bivariate & Multivariate Analysis

Bivariate and multivariate analysis was conducted to see the comparison and relationship between two or more data. At this stage a correlation table was made to determine the correlation of the data so that it is found that Age on Education has a slightly positive correlation and Income on education also has a slightly positive correlation shown in Figure 3.

Figure 3. Correlation Table.

From the results of the comparison between data at this stage, it was also found that customers with low education with age above 40 and with a income of less than 1.5 million have potention to became regular customers, while those in the age range 30-35 had the highest percentage of ages who had purchased more than once.

2.4 Model Building and Evaluation

2.4.1 Decision Tree Model

This research will use the application of machine learning to build model. Machine Learning itself is one of the techniques in Artificial Intelligence that could learn patterns from a dataset [19]. Meanwhile, the machine learning technique used in this research is a supervised learning with classification model using Decision Tree algorithm. In general, the Machine Learning model that will be used in this research can be seen in Figure 4.

At this stage the 5000 data that are ready to use (dataset) are divided into two groups, which are training data and testing data with a proportion of 70% training data and 30% testing data. For each data distribution a model is made using the Decision Tree classification method, then for parameter setting the selected criteria are using gain ratio or information gain. If the frequency of the potential class is 10% and the frequency of the not potential class is 90% in label attribute, then the not potential class will become the dominant class, the tree will be biased towards the dominant class. Due to data imbalance, it is necessary to add weight.

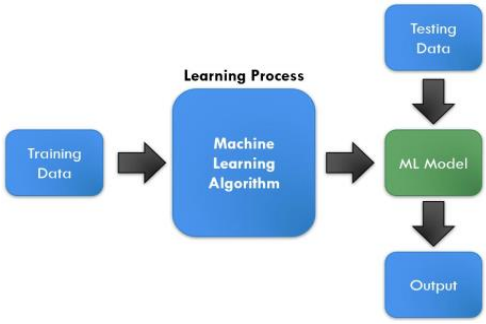

Figure 4. Machine Learning Model

2.4.2 CART (Classification and Regression Tree)

CART is a recursive partitioning method used for regression and classification. CART is constructed by solving a subset of the dataset using a predictor variable to create two child nodes repeatedly, starting from the entire dataset. The aim is to produce a subset of data that is as homogeneous as possible to classify the target variables [20].

At the start of the process, a training set consisting of confidential notes should be available. a training tool used to construct a tree that allows assigning classes to target new record variables based on the value of another variable or independent variable [21].

CART builds a binary tree by dividing the records at each node based on a function of the input variables. The first task to run is to determine the independent variable which is the best spliiter. The best splitter is the vertex diversity of the derived divisors. No more splitting nodes called leaf nodes [21].

Breaking records at each node causes the number of records to decrease from the root node to the child node to the leaf node. The fewer the quantity records, the less representative the node is. The result is that the tree model can only accurately predict records in the training set but cannot predict new records from outside the training set accurately or overtraining. To reduce overtraining can be done by tuning, if tuning can't handle overtraining, then pruning is used which produces many candidate subtrees [21].

Several candidate subtrees are selected based on their ability to predict new records. Selection requires a new dataset, a test set containing new records that are different from the notes in the training set. Each candidate subtree is used to predict the records in the test set. The subtree that gives the smallest error is selected as the model tree [21].

The final step is to evaluate the selected subtree by applying it to a new dataset which is the validation set.

DOI: https://doi.org/10.29207/resti.v7i5.4531

The error value obtained from the validation set is used to predict the expected performance prediction model [21].

3. Results and Discussions

3.1 Analysis of CART Decision Tree Model

From the findings of the initial data, there is a data imbalance, so it is necessary to add weight. There is confusion matrix generated based on model that produced shows that it is still overfitting shown in Figure 5.

```
Accuracy : Train : 0.9094285714285715 Test: 0.7773333333333333
Recall   : Train : 0.9910714285714286 Test: 0.18055555555555555
```

Figure 5. Confusion Matrik Before Tuning

Because the model is still overfitting, where the accuracy and recall values of the training data is still far from the accuracy and recall value of the testing data, then the tuning process is carried out with a grid search. Grid search is a way to find the best parameters used for modeling in machine learning. If we use this grid search, we can find out which hyperparameter is the best that we want to use for modeling in a machine learning algorithm. In this step parameters from the model are used to perform validation for each combination of models and hyperparameters automatically. In this process, the best parameter generated are hyperparameter with max depth = 6, max leaf nodes for tree = 20, and min samples leaf for tree = 7. The result showed that tuning process can overcome overfitting, and the model has shown fit results which is proven by the confusion matrix in Figure 6.

```
Accuracy : Train : 0.9045714285714286 Test: 0.9
Recall   : Train : 0.011904761904761904 Test: 0.006944444444444444
```

Figure 6. Confusion Matrik After Tuning

In figure 6 the result showed that accuracy and recall values of the training data already not far from accuracy and recall values from data testing, so the model can be said to be fit and can be used to help decision making. From an already made model, it is estimated that important features include education with the scope of bachelor and diploma degree, age with scope older than 25 years old, income with a salary scope greater than equal 5 million, then followed by occupation, that belong to jobless, teaching staff and private sector as shown in Figure 7.

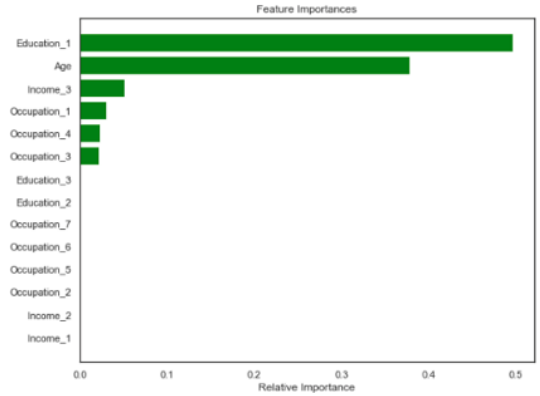

Figure 7. Important Features

As for the decision tree model that successfully generated after the tuning process and model can be said already fit is shown in Figure 8

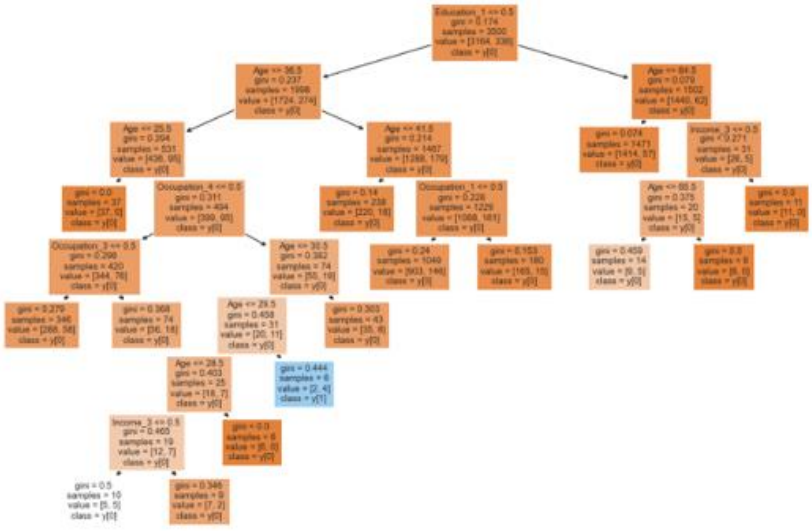

Figure 7. Decision Tree Model After tuning

From Figure 8 then several important rules that used for helping decision making base on decision tree above for determining potential customers shown as follows:

R1 : IF **Education** <= high school level and **Age** <= 37 **then** not potential customer

DOI: https://doi.org/10.29207/resti.v7i5.4531

R2 : IF **Education** <= high school level and **Age** <= 37 and **occupation** = teacher and **income** <= 5 million rupiah **then** not potential customer

R3 : IF **Education** <= high school level and **Age** <= 42 **then** not potential customer

R4 : IF **Education** <= high school level and **Age** <= 65 **then** not potential customer

R4 : IF **Education** <= high school level and **Age** <= 42 and **occupation** = jobless **then** not potential customer

R5 : IF **Education** <= high school level and **Age** <= 65 and **income** <= 5 million rupiah **then** not potential customer

R6 : IF **Education** <= high school level and **Age** <= 66 and **income** <= 5 million rupiah and **then** not potential customer

R7 : IF **Education** <= high school level and **Age** <= 37 and **occupation** = teacher or **occupation** = enterpreneur **then** potential customer

After rules successfully generated then the next process is to checking tuning process result. When tuning process if the model still not fit then the next step process which is prunning will be executed. Because the model is already fit, there no need to do pruning process and the model can be applied in decision making. Based on the model and analysis it was found that the best feature that sequentially from Education, Age, Income and the last one is occupation. From the main features the right marketing programs can be focused or aimed on people with high school level education background and below, and with a second priority people with occupation teacher or enterpreneur and also Age below 37 years old.

3.2 Analysis of Precision Marketing Implementation

To test the effectiveness of precision marketing on marketing program of orebae.com e-commerce with model that already developed, the model was applied by applying the rules generated from the decision tree on marketing strategy and also based on the important features that have been generated on monthly sales. The test will be focused on how much volume and marketing costs for the month of July which does not apply the precision marketing model and August which has used the precision marketing model and then the results will be compared. The results are shown in Table 4, where thanks to precise marketing, the monthly sales volume increased after implementing model, which is 32.65% higher than before and for marketing cost is 50%. lower than last month. The results of the study show precision marketing has good impact to sales volume and marketing cost.

After implementing precision marketing model, marketing cost of orebae.com is significantly reduced. This is because orebae.com only carries out the marketing plan for the designated customers, which largely avoids the invalid marketing plan and reduces the marketing cost.

Table 4. Comparison of Model Results

Times	July, 2022	August, 2022
Monthly sales volume/n	267	392
Marketing cost/jt	15	7.5

4. Conclusion

In this research, the decision tree algorithm can be used and applied to create new model of precision marketing. Taking the SME e-commerce orebae.com as an example, a decision tree was constructed with classification rules based on the analysis of customer information collected and yielded satisfactory results. Finally, categories of customers with high purchase probability are obtained. This customer category then becomes the main target of the marketing program. From analysis process and rules that generated, show that the best target for marketing campain are customer with education background high school level education and below, occupation teacher or enterpreneur and also Age below 37 years old. Finally, after adjust marketing campain based on model result, it was found that the company's sales volume increased significantly, and marketing costs decreased, indicating that the built decision tree is quite effective and reliable for precision marketing.

Suggestions for further research, it is highly recommended to be able to combine it with mining data from user interactions when using e-commerce such as when choosing a product, how long to look at a product, what product categories are often viewed etc. So that precision marketing can be more precise in reaching the expected customer categories.

Acknowledgment

Thanks to orebae.com for being ready providing data for research and willing to collaborate in research completion.

References

[1] Ö. Aydın, A. Ozen, F. N. Gürel, and D. Mhlanga, "THE IMPACTS OF DIGITAL TRANSFORMATION Business environments in Bulgaria and Brazil View project Competitiveness View project," 2020. [Online]. Available: https://www.researchgate.net/publication/344071634

[2] D. Tanna, D. Raval, and Z. Raval, "Internet Marketing Over Traditional Marketing," 2014. [Online]. Available: https://www.researchgate.net/publication/267395144

[3] J. Cheng, "An evaluation strategy for commercial precision marketing based on artificial neural network," *Revue d'Intelligence Artificielle*, vol. 34, no. 5, pp. 637–644, 2020, doi: 10.18280/ria.340515.

[4] I. Maryani and D. Riana, "Clustering and profiling of customers using RFM for customer relationship management recommendations," *2017 5th International Conference on*

DOI: https://doi.org/10.29207/resti.v7i5.4531

Cyber and IT Service Management, CITSM 2017, pp. 2–7, 2017, doi: 10.1109/CITSM.2017.8089258.

[5] Z. Zhu, Y. Zhou, X. Deng, and X. Wang, "A graph-oriented model for hierarchical user interest in precision social marketing," *Electron Commer Res Appl*, vol. 35, May 2019, doi: 10.1016/j.elerap.2019.100845.

[6] M. jie Liao, J. Zhang, R. mei Wang, and L. Qi, "Simulation research on online marketing strategies of branded agricultural products based on the difference in opinion leader attitudes," *Information Processing in Agriculture*, vol. 8, no. 4, pp. 528–536, Dec. 2021, doi: 10.1016/j.inpa.2020.12.001.

[7] F. A. Shaw, X. Wang, P. L. Mokhtarian, and K. E. Watkins, "Supplementing transportation data sources with targeted marketing data: Applications, integration, and internal validation," *Transp Res Part A Policy Pract*, vol. 149, pp. 150–169, Jul. 2021, doi: 10.1016/j.tra.2021.04.021.

[8] J. R. Saura, "Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics," *Journal of Innovation and Knowledge*, vol. 6, no. 2, pp. 92–102, Apr. 2021, doi: 10.1016/j.jik.2020.08.001.

[9] L. I. U. Xiao-Yuan, "Agricultural products intelligent marketing technology innovation in big data era," in *Procedia Computer Science*, 2021, vol. 183, pp. 648–654. doi: 10.1016/j.procs.2021.02.110.

[10] W. Jun *et al.*, "Evaluation of precision marketing effectiveness of community e-commerce–An AISAS based model," *Sustainable Operations and Computers*, vol. 2, no. April, pp. 200–205, 2021, doi: 10.1016/j.susoc.2021.07.007.

[11] Z. You, Y. W. Si, D. Zhang, X. Zeng, S. C. H. Leung, and T. Li, "A decision-making framework for precision marketing," *Expert Syst Appl*, vol. 42, no. 7, pp. 3357–3367, 2015, doi: 10.1016/j.eswa.2014.12.022.

[12] G. Zhu, "Precision Retail Marketing Strategy Based on Digital Marketing Model," *Science Journal of Business and Management*, vol. 7, no. 1, p. 33, 2019, doi: 10.11648/j.sjbm.20190701.15.

[13] Y. Zheng, "Decision tree algorithm for precision marketing via network channel," *Computer Systems Science and Engineering*, vol. 35, no. 4, pp. 293–298, 2020, doi: 10.32604/csse.2020.35.293.

[14] H. Wang, J. Wang, and Z. Zhong, "Research on precision marketing strategy based on cluster analysis algorithm," *Proceedings - 2020 International Conference on E-Commerce and Internet Technology, ECIT 2020*, pp. 208–211, 2020, doi: 10.1109/ECIT50008.2020.00054.

[15] S. Budilaksono *et al.*, "Customer Profilling for Precision Marketing using RFM Method, K-MEANS algorithm and Decision Tree," *SinkrOn*, vol. 6, no. 1, pp. 191–200, 2021, doi: 10.33395/sinkron.v6i1.11225.

[16] Z. Zhu, L. Kong, X. Deng, and B. Tan, "A 2020 perspective on 'A graph-oriented model for hierarchical user interest in precision social marketing,'" *Electron Commer Res Appl*, vol. 41, no. February, p. 100962, 2020, doi: 10.1016/j.elerap.2020.100962.

[17] Y. Xiao and F. Ling, "On E-Commerce Precision Marketing Strategy Based on Big Data," *Big Data and Cloud Innovation*, vol. 3, no. 1, pp. 36–39, 2019, doi: 10.18063/bdci.v3i1.1149.

[18] B. Zhang and B. Zhang, "Precise marketing of precision marketing value chain process on the H group line based on big data," *Journal of Intelligent and Fuzzy Systems*, vol. 35, no. 3, pp. 2837–2845, 2018, doi: 10.3233/JIFS-169637.

[19] H.-D. Wehle, "Machine Learning, Deep Learning, and AI: What's the Difference?" 2017. [Online]. Available: https://www.researchgate.net/publication/318900216

[20] N. Kappelhof *et al.*, "Evolutionary algorithms and decision trees for predicting poor outcome after endovascular treatment for acute ischemic stroke," *Comput Biol Med*, vol. 133, Jun. 2021, doi: 10.1016/j.compbiomed.2021.104414.

[21] D. H. Lee, S. H. Kim, and K. J. Kim, "Multistage MR-CART: Multiresponse optimization in a multistage process using a classification and regression tree method," *Comput Ind Eng*, vol. 159, Sep. 2021, doi: 10.1016/j.cie.2021.107513.

DOI: https://doi.org/10.29207/resti.v7i5.4531