Predictive Buyer Behavior Model as Customer Retention Optimization Strategy in E-commerce

Muhammad A. A. Hakim[1] and Terttiaavini[2]
[1] Digital Business, Faculty of Business and Social Sciences, Binawan University, Indonesia
[2] Computer Science, Faculty of Computer Science and Sciences, Indo Global Mandiri University, Indonesia

Corresponding author: Muhammad A. A. Hakim (e-mail: muhammadariv@gmail.com)

ABSTRACT Lazada is one of the rapidly growing E-commerce platforms in this digital era. One of the main challenges faced by Lazada is customer retention, where customers make purchases once or a few times before switching to other platforms. Therefore, it is important to understand buyer behavior in E-commerce through customer prediction to identify factors influencing retention. This study employs the Random Forest (RF) method to analyze Lazada customer data and formulate more effective marketing strategies. The analysis is conducted by loading preprocessed datasets into the KNIME workflow and utilizing various nodes and algorithms available in KNIME to build and evaluate predictive models. The Random Forest model is trained multiple times to achieve the highest Accuracy rate, which is 72.472%, with a fairly high level of agreement and a balanced trade-off between recall and precision. Additionally, this model successfully predicts that customers purchasing electronic equipment are potentially churning at a rate of 3.85%. Subsequently, customer strategy analysis for customer retention optimization in the E-commerce industry is conducted through data visualization using Tableau. Predictive analysis of customer behavior serves as a strong foundation for formulating effective retention strategies in the E-commerce industry. With this approach, Lazada can enhance customer experience and ensure sustainability in facing the increasingly fierce competition in the digital market.

KEYWORDS Customer Retention Optimization Strategy, E-Commerce, Predictive Behavior Model, Random Forest

I. INTRODUCTION

In the current digital era, the e-commerce industry, such as Lazada, has emerged as one of the rapidly growing digital platforms [1]. One of the main challenges faced by the Lazada platform is customer retention. Customer retention refers to the desire of customers to make repeat purchases online [2]. It has become the primary focus for Lazada due to the high cost of customer acquisition and the long-term benefits that can be derived from loyal customers. Building a strong customer base on the E-commerce platform can be achieved by increasing the Lifetime Value (LTV) of customers, leveraging recommendations and positive reviews, offering incentives for repeat purchases, and enhancing customer loyalty through optimal experiences. These strategies are key to retaining customers in the long run [3].

Although Lazada offers convenience in shopping, customer retention remains the primary focus, considering that most of them only make purchases once or a few times before eventually switching to other platforms. Therefore, it is important to understand buyer behavior in E-commerce to identify factors influencing customer retention. The factors influencing customer retention are closely related to predictive methods. In predictive methods, these factors are used as features to create models that can predict whether a customer is likely to stay or churn. By understanding the factors affecting retention, predictive models can be better trained to identify behavioral patterns that lead to customer retention, thus enabling E-commerce platforms to take more proactive steps in retaining customers.

Customer prediction allows Lazada to group buyers based on shopping loyalty patterns, devise more effective marketing strategies, and enhance the overall shopping experience.

This study employs the Random Forest (RF) method due to its ability to handle complexity and noise in data, as well as its capability to address overfitting issues. By constructing multiple decision trees randomly, RF can generate more stable and accurate predictions compared to other predictive methods [4][5]. Previous research [6] addressed the same topic, focusing on consumer review classification using Random Forest and SMOTE. They reported an Accuracy rate of 75%, which increased to 77% when utilizing 8000 max_features [6]. Another study [7] aimed to predict customer churn in a campus fashion company by identifying customers who had not made transactions for more than 365 days as 'churned'. In this research, the Random Forest Classifier model achieved the highest Accuracy compared to three other Machine Learning models. The analysis results indicated that the customer churn rate was 24.54%, while the non-churned customers accounted for 75.46%. The top five customers originated from Jakarta, West Java, Central Java, East Java, and Yogyakarta provinces, with the highest total transaction value reaching 3,997,936,774 [7]. Furthermore, research [8] performed a descriptive analysis of Shopee user data. The results revealed that more users discontinued using Shopee (54.8%) compared to those who remained active (45.2%). The classification model utilized was Random Forest due to its superior performance [8].

Based on the findings of these studies, the Random Forest method has proven to be effective in addressing the challenge of customer retention in e-commerce platforms. With its demonstrated ability to generate accurate predictions regarding customer behavior, such as churn prediction or customer segmentation [5][9], Random Forest can assist platforms like Lazada in devising more targeted marketing strategies and enhancing the shopping experience for customers [10]. Furthermore, this research also identifies customer segments vulnerable to churn, enabling proactive efforts to minimize the rate of customers leaving the platform [11]. With a better understanding of customer behavior and factors influencing churn, Lazada can take appropriate steps to improve customer retention among those vulnerable to churn and enhance the overall user experience [12].

Based on the aforementioned background, the objectives of this research are: 1) To predict customer behavior on e-commerce platforms, particularly Lazada, using the Random Forest method, and 2) To design more effective marketing strategies based on these prediction results, aiming to enhance customer retention and ensure business sustainability in this competitive digital era. Thus, this research will provide a significant contribution to managing customer relationships and improving customer retention on e-commerce platforms.

II. LITERATURE STUDY

A. RANDOM FOREST (RF)

The Random Forest method is one of the techniques in data analysis used to predict or classify data. This technique works by dividing the dataset into many small subsets and then building decision trees for each of these subsets. Subsequently, the results from all these decision trees are combined or averaged to produce more accurate predictions. This method is commonly used in machine learning due to its ability to address overfitting and provide stable results in various situations. Random Forest is an ensemble algorithm that utilizes the concept of decision trees in its model formation. Although the Random Forest algorithm itself does not directly use entropy, the decision trees it employs can utilize entropy as one of the criteria for node splitting when constructing the tree. Equation (1) represents the formula for calculating entropy where Y is the set of cases and p(c | Y) represents the proportion of class c values in Y.

$$Entropy\ (Y) = -\sum_{i} \mathrm{p(c|Y)} \log_2 \mathrm{p(c|Y)} \tag{1}$$

$$Information\ Gain\ (Y,a) = Entropy\ (Y) - \sum_{ve} values\ \frac{Y_v}{Y_a}\ Entropy\ (Y_v) \tag{2}$$

Equation (2) is the formula for calculating information gain where values $\frac{Y_v}{Y_a}$ a represent all possible values in the set of cases a, Y_v is the subclass of Y with class v related to class a. Y_a is all the values corresponding to a.
The selection of attributes to be used as nodes, either as roots or internal nodes, depends on the highest information gain value possessed by the available attributes. Information Gain is used to determine the most beneficial attribute in decision-making. The value of Information Gain can be found using the formula in (3), and the gain ratio value can be observed in (4) , where Split Information $Split\ Information\ (S,A)$ is the estimated entropy value of the input variable S which has class c, while $\frac{|S_i|}{|S|}$ represents the probability of class i within that attribute.

$$Split\ Information\ (S,A) = \sum_{i}^{c} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \tag{3}$$

$$Gain\ ratio\ (S,A) = \frac{Information\ (S,A)}{Split\ information\ (S,A)} \tag{4}$$

B. CHURN ANALYSIS

Churn analysis is the process of understanding, identifying, and managing the behavior of customers or users who cease using a product or service. It is a crucial aspect of customer relationship management (CRM) and customer retention strategies. Here are the common steps in churn analysis:

1) CUSTOMER DATA COLLECTION

The initial step in churn analysis is gathering relevant customer data. This data may include information such as customer profiles, transaction history, interactions with products or services, and more.

2) DATA EXPLORATION

Once the data is collected, the next step is to explore the data to understand patterns and trends that may be associated with churn behavior. This involves descriptive statistical analysis, data visualization, and identification of features that may influence churn.

3) PREDICTIVE MODELING

One key aspect of churn analysis is building predictive models to forecast future churn behavior. This involves using machine learning techniques such as Logistic Regression, Decision trees, Random Forests, or neural networks. These models utilize historical customer data to predict the probability of churn for new or existing customers.

4) MODEL VALIDATION

After building the predictive model, the next step is to test and validate its performance using independent data. This is important to ensure that the model has good predictive capability for accurately forecasting churn behavior.

5) CUSTOMER SEGMENTATION

One common strategy used in churn management is to divide customers into smaller groups based on similar characteristics. This is called customer segmentation. This segmentation helps companies better understand customer behavior and design more effective retention strategies.

6) IMPLEMENTATION OF RETENTION STRATEGIES

Based on churn analysis results, companies can design and implement appropriate customer retention strategies. This may involve improving the customer experience, offering incentives, or loyalty programs to encourage customers to continue using products or services.

7) MONITORING AND EVALUATION

Lastly, it is important to continually monitor and evaluate the effectiveness of customer retention strategies. This allows companies to adjust and improve their strategies over time in response to changes in customer behavior and the business environment.

Churn analysis is a continuous and iterative process. By understanding the factors influencing churn and designing appropriate strategies, companies can reduce churn rates and improve customer retention, which in turn can contribute to the long-term growth and success of the company.

III. RESEARCH METHODOLOGY

The research methodology in this study comprises several stages that outline the process of data collection and analysis, as well as the development of customer retention strategies. The research stages are described in Figure 1 below.

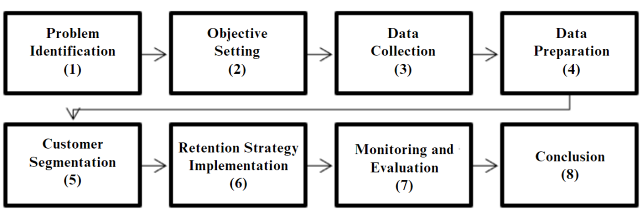

Figure 1. Research Stages

The stages are described in detail as follows:

1) DATA COLLECTION

The initial stage involves collecting transaction data, product preferences, and shopping behaviour from Lazada customers. This data can be obtained from Lazada's internal database or through customer survey methods. The data is obtained from sales transactions on Lazada in 2022. In this study, the focus is solely on the sales of electronic goods because the main objective is to understand customer behaviour in the context of purchasing electronic products specifically. The sales data on Lazada in 2022 is displayed in table 1.

TABLE I
ELECTRONIC SALES DATA ON LAZADA IN 2022

Dataset	Range / Frekuensi	Persentase
Catagory		
Harddisk-eksternal	4422	40,41
Laptop	701	6,41
Smart-tv	1290	11,79
China OEM	2	0,02
flash-drives	3318	30,32
televisi-digital	1211	11,07
BrandName		
No Brand	943	8,57
Merk Dll	1-20 / 920	8,37
Akari, Aoyama, Aqua, Bestrunner, Carcool, Hisense, HP COMPAQ, Ichiko, Import, Led Coocaa, Microsoft, MSI, Multi, Niko, OEM, Universal_Brand, Vandisk, Xiaomi	20-60 / 565	5,14
Apple, Changhong, Flashdisk, Ikedo, Orico, Philips, Sony,TCL,Vakind,VGen	51-100 / 707	6,43
Adata, Coocaa, Kingston, Panasonic, Transcend, Universal	101-200 / 852	7,75
Dell, LG, Seagate, Sharp, WD	201-300 / 1121	10,19
Polytron, Samsung	301-400 / 656	5,96
Acer, China OEM	401-500 / 864	7,86
HP	501-600 / 582	5,29
Toshiba	601-700 / 674	6,13
Asus, Lenovo	900-1000 / 1856	16,88
SanDisk	1000-1300 / 1258	11,44
TotalReviews		
Aoyama, AVITA, DBest, E-link, Hisense, Maxtor, Merk Lainnya, Microsoft, NYK, OEM, OneGood, OTG, PX, Robot, Sanyo, SelaluAda, SP, VANDISK, YYSL	51-100 / 1.399	0,47
Bmstore, Boneka_Nizza, Casing, Hardcase, Multi, Qflash, Redcolourful, Trisonic, Universal_Brand, Universally	101-150 / 1.186	0,40
Akari, Best CT, Bestrunner, Carcool, Hiqueen, JvGood, Sony, Vitron	151-200 / 1.424	0,48
Apple, EsoGoal, KLEVV, Max, M-Tech, Niko,Rendys chem, V-Gen	201-300 / 1.944	0,65
Ikedo, Lazada, Vakind	301-400 / 1.010	0,34
Trend's, Universal Indonesia	401-500 / 870	0,29
Aqua, Dell, Lexar, SS,Transcend	500-1000 / 3.985	1,33
Acer, Changhong, Good Shop, UGREEN	1001-1500 / 4.468	1,49
China OEM, Kingston, Panasonic, Universal, WARM	1501-2000 / 8.964	3,00
Ichiko, Orico, Seagate	2001-2500 / 6.525	2,18
Flashdisk, WD	2501-3000 / 5.471	1,83
Adata, LG	3001-3500 / 6.289	2,10
TCL	3500-4000 / 4.232	1,42

Lenovo, Philips, Polytron, Toshiba	5000-10000 / 30.079	10,06
Asus, Coocaa, Samsung, SanDisk, Sharp, Xiaomi	10000-70000 / 210.586	70,44
No brand	1.183	0,40

2) DATA PREPARATION (PRE-PROCESSING DATA)

The collected data will be prepared for further analysis, including data cleaning to remove anomalies or inconsistencies, as well as data processing to prepare it into a suitable format for analysis. In the data preparation stage, the processes include data cleaning, data processing, selection of the most relevant variables, and data validation. The data is obtained from sales on Lazada, which includes various variables such as itemId, Category, Name, BrandName, url, Price, AverageRating, TotalReviews, and RetrievedDate. From this data, the most relevant variables to be used in cluster analysis are itemId, Price, AverageRating, and TotalReviews.

3) PREDICTION ANALYSIS

Prediction analysis using the Random Forest model involves selecting, training, validating, and testing the model to make predictions on new data [13]. After validation and testing, the prediction results are evaluated using performance metrics such as Accuracy, Precision, Recall, and F1-score, while considering the interpretation of the results to understand the factors influencing the model's predictions.

In this study, the preprocessed data resulted in a cleaned dataset free from missing values, outliers, and duplicates. Additionally, relevant features have been extracted or processed, and the data has been transformed or normalized to ensure consistency and uniformity.

Prediction analysis is conducted using the KNIME application [14][15]. The analysis steps begin by loading the preprocessed dataset into the KNIME workflow. Various nodes and algorithms available in KNIME are then used to build and evaluate predictive models. The first step involves using the Excel Reader node and Column Filter to filter the data, including only relevant columns for prediction analysis, such as itemId, Price, AverageRating, and TotalReviews. The data is then partitioned using the Partitioning node, with a training dataset of 85% and a testing dataset of 15%. After that, the model is learned using the Random Forest Learner node, and tested using the Random Forest Predictor node with the test dataset. The prediction results are evaluated using appropriate evaluation metrics and can be viewed through the Score and Table View nodes. Figure 2 shows the workflow diagram of the prediction model using Random Forest (RF).

The Random Forest model was trained repeatedly with various parameter variations to achieve the highest Accuracy score. After several training sessions, the following results were obtained: Accuracy score = 72.472%; Cohen's Kappa = 0.691%; correct classified = 10524; wrong classified = 422; Recall score = 0.85; Precision score = 0.95, and F1-Score = 0.785.

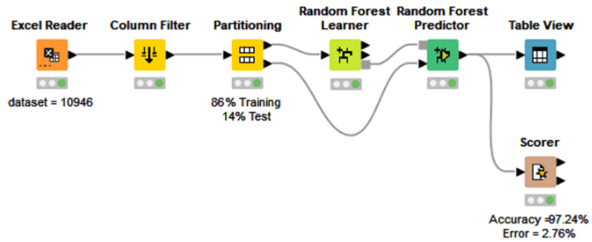

Figure 2. Workflow diagram for prediction model using Random Forest (RF)

The conclusions drawn from these testing results are as follows:

1. Accuracy score: The model has an Accuracy of 72.472%, indicating the percentage of correct predictions out of the total predictions made.
2. Cohen's Kappa: The Cohen's Kappa value is 0.691%, indicating the level of agreement between the model's predictions and the expected predictions, after correcting for chance agreement.
3. Correct classified: A total of 10524 samples were correctly classified by the model.
4. Wrong classified: There were 422 samples classified incorrectly by the model.
5. Recall score: The model's sensitivity, or recall score, is 0.85, indicating the model's ability to identify a large number of true positive cases.
6. Precision score: The model's precision rate is 0.95, depicting the proportion of true positive outcomes among all outcomes predicted positively by the model.
7. F1-Score: The harmonic mean of recall and precision, or F1-Score, is 0.785. This provides an overall overview of the model's performance by considering the balance between recall and precision.

Overall, the model demonstrates a fairly good performance with decent Accuracy, a good balance between recall and precision, and a relatively high level of agreement with Cohen's Kappa at 0.691%. The prediction of churn customers is 422 customers.

4) CUSTOMER SEGMENTATION

One common strategy used in managing churn is to divide customers into smaller groups based on similar characteristics, known as customer segmentation. This segmentation helps companies better understand customer behavior and design more effective retention strategies. Analysis can be visualized using Tableau, enabling researchers to quickly observe patterns and trends in customer data [16][17]. Customer analysis is divided into several segments:

1. Sales Analysis Based on Number of Sales per Category

   Sales visualization based on the number of sales per category is presented in Figure 3.

Category	
beli-harddisk-eksternal	18.289
beli-laptop	3.010
beli-smart-tv	5.522
China OEM	87.480
jual-flash-drives	13.606
No Brand	87.480
shop-televisi-digital	5.200

Figure 3. **Sales Analysis Based on Number of Sales Per Category**

Based on product sales data, China OEM and No Brand show significant popularity with each reaching 87,480 units, while Buy External Hard Drives, Sell Flash Drives, and Buy Smart TVs have relatively high sales with 18,289, 13,606, and 5,522 units respectively. On the other hand, sales of Buy Laptops (3,010 units) and Shop Digital Televisions (5,200 units) are relatively lower. This analysis indicates that customer management strategies should emphasize maintaining the popularity of the most favored products while strengthening sales of less popular products by launching special promotions or enhancing customer service quality.

2. Sales Analysis Based on Category
Sales visualization based on category is presented in Figure 4.

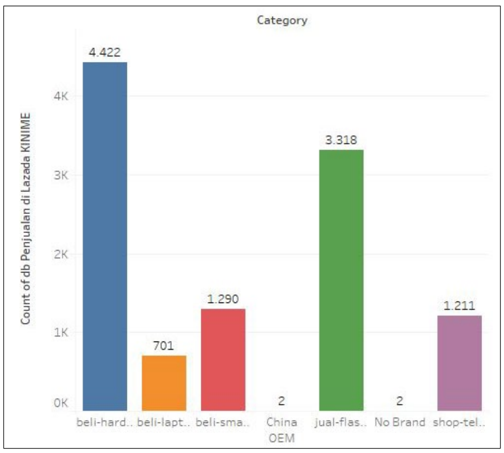

Figure 4. **Sales Analysis Based on Category**

China OEM leads the sales with 87,480 units, while Buy Laptops records the lowest sales with only 3,010 units. An effective customer management strategy should consider these differences, focusing on strengthening product availability and improving customer service quality for widely favored products, as well as developing more aggressive and innovative marketing strategies to increase interest and loyalty towards products with lower sales.

3. Sales Analysis Based on Brand Name
Sales visualization based on brand name is presented in Figure 5.

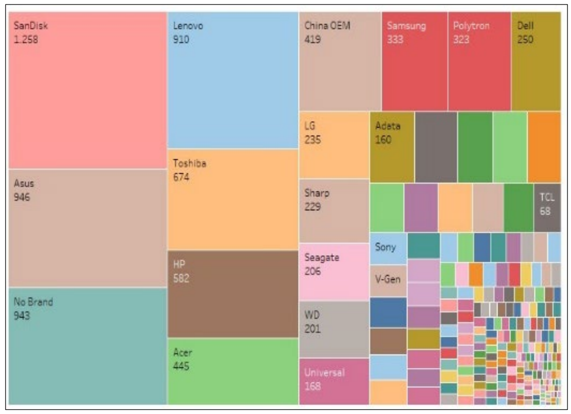

Figure 5. **Sales Analysis Based on Brand Name**

Based on sales data by BrandName, it can be seen that most sales are dominated by several major brands, such as Asus, Coocaa, Samsung, SanDisk, Sharp, and Xiaomi, which account for a total sales of 210,586 units or approximately 70.44% of total sales. Meanwhile, some other brands have lower contributions to sales, such as Apple, Dell, and LG, each having sales ranging from 201-300 to 3001-3500 units. Customer management strategies derived from this data include focusing on major brands dominating sales by improving product availability, providing special promotions, and enhancing customer service. On the other hand, attention should also be given to brands with lower sales by evaluating customer needs and preferences, as well as possibly developing more creative and targeted marketing strategies to increase brand interest and awareness.

4. Sales Analysis Based on Product Reviews
Sales visualization based on product reviews is presented in Figure 6.

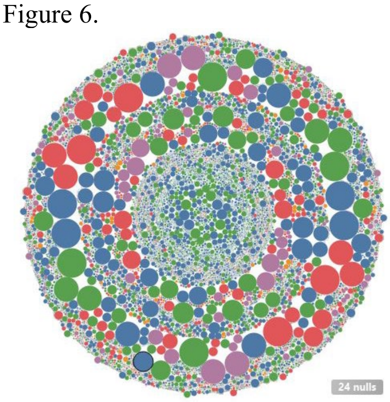

Figure 6. **Sales Analysis Based on Product Reviews**

Based on TotalReviews data, it can be observed that most brands have a diverse range of review counts. However, most brands have review counts concentrated in lower ranges, with approximately 70.44% of major brands like Asus, Coocaa, Samsung, SanDisk, Sharp, and Xiaomi having over 10,000 reviews. On the other hand, most other brands have review counts below 2,000. Customer management strategies can focus on brands with low review counts by increasing customer engagement, encouraging product reviews, and enhancing brand visibility. For brands with high review counts, it is important to maintain and strengthen relationships with loyal customers, encourage positive interactions, and provide prompt and satisfactory responses to reviews given. This will help increase customer trust and expand the loyal customer base.

5) IMPLEMENTATION OF RETENTION STRATEGIES

Based on the analysis of the four segments, the implementation of customer retention strategies can be carried out with a comprehensive approach. First, to improve customer retention based on sales per category, strategies can focus on maintaining the popularity of the most favored products such as China OEM and No Brand, while also strengthening the sales of less popular products like Buy Laptops and Shop Digital Televisions by launching special promotions or enhancing customer service quality. Second, considering the analysis of sales based on categories, retention strategies can focus on strengthening product availability and improving customer service quality for widely favored products, as well as developing more aggressive and innovative marketing strategies for products with low sales. Third, considering sales based on brands, retention strategies can focus on major brands dominating sales by improving product availability, providing special promotions, and enhancing customer service, while also paying special attention to brands with lower sales through evaluation of customer needs and preferences and development of more creative and targeted marketing strategies. Fourth, considering the analysis of sales based on product reviews, retention strategies can focus on increasing customer engagement and encouraging product reviews for brands with low review counts, while also maintaining and strengthening relationships with loyal customers and providing satisfactory responses to reviews for brands with high review counts. With a holistic and focused approach, the implementation of these retention strategies is expected to increase customer loyalty and strengthen the company's position in the e-commerce market. Monitoring and Evaluation. Finally, it is important to continuously monitor and evaluate the effectiveness of customer retention strategies. This allows the company to adjust and improve its strategies over time according to changes in customer behavior and the business environment.

IV. CONCLUSION

The findings of this research are as follows:

1. Predictive Analysis with Random Forest resulted in an accuracy rate of 72.472%, as well as a good balance between recall and precision. The test results indicate that the model performs well in predicting customer behavior based on factors influencing retention. The Random Forest method is capable of identifying customer segments vulnerable to churn, enabling proactive efforts to minimize the rate of customer attrition from the platform.
2. The Retention/Churn analysis shows that Lazada still faces challenges in maintaining the popularity of electronic products, where a large percentage of loyal customers accounts for 96.15%, while customers indicated as churning are 3.85%.
3. In enhancing customer retention on E-commerce platforms like Lazada, several key strategies can be employed. Firstly, by segmenting customers based on their shopping behavior, companies can better understand their needs to devise more effective marketing strategies. Then, by providing personalized experiences through product recommendations and relevant promotional offers, relationships with customers can be strengthened. Additionally, improving product availability and quality, as well as innovation in marketing, are also important to attract customer interest and maintain their loyalty. Continuous monitoring and evaluation of retention strategies are also necessary to adjust the company's approach to changes in customer behavior and market trends. By implementing these strategies comprehensively, Lazada can strengthen its position in the E-commerce market and sustain business growth.

AUTHORS CONTRIBUTION

Muhammad Arif Abdul Hakim: Investigation, data collection, analysis, review writing, and editing.
Terttiaavini: Data analysis, validation, Interpretation of results, visualization.

COPYRIGHT

REFERENCES

[1] M. A. A. Hakim, "Analysis Of The Influence Of Content Personalisation On Consumer Purchasing Decisions In The E-commerce Industry," *J. MIRTE*, vol. 2, no. 2, pp. 10–11, 2023.
[2] M. I. Eid, "Determinants of E-commerce customer satisfaction, trust, and loyalty in Saudi Arabia," *J. Electron. Commer. Res.*, vol. 12, no. 1, pp. 78–93, 2011.
[3] M. Harahap, Y. Lubis, and Z. Situmorang, "Data Science bidang Pemasaran : Analisis Prilaku Pelanggan," *Data Sci. Indonesai*, 2021.
[4] N. F. Sahamony, T. Terttiaavini, and H. Rianto, "Analisis Perbandingan Kinerja Model Machine Learning untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 413–422, 2024, [Online]. Available: https://journal.irpi.or.id/index.php/malcom/article/view/1210
[5] I. P. Putri, T. Terttiaavini, and N. Arminarahmah, "Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 257–265, 2024, doi: 10.57152/malcom.v4i1.1078.
[6] N. Istiqamah and M. Rijal, "Klasifikasi Ulasan Konsumen Menggunakan *Random Forest* dan SMOTE," *J. Syst. Comput. Eng.*, vol. 5, no. 1, pp. 66–77, 2024, doi: 10.61628/jsce.v5i1.1061.

[7] T. N. Muthmainnah and A. Voutama, "Pendekatan Data Science Untuk Menemukan Customer *Churn* Pada Perusahaan Fashion Dengan Metode Machine Learning," *J. Teknol. Sist. Inf. dan Sist. Komput. TGD*, vol. 6, no. 2, pp. 463–471, 2023, [Online]. Available: https://ojs.trigunadharma.ac.id/index.php/jsk/index

[8] W. Fajrin Mustafa, S. Hidayat, and D. Hatta Fudholi, "Prediksi Retensi Pengguna Baru Shopee Menggunakan Machine Learning," *J. Media Inform. Budidarma* , vol. 8, no. 1, pp. 612–623, 2024, doi: 10.30865/mib.v8i1.7074.

[9] N. F. Sahamony, T. Terttiaavini, and H. Rianto, "Analysis of Performance Comparison of Machine Learning Models for Predicting Stunting Risk in Children ' s Growth," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. April, pp. 413–422, 2024.

[10] A. A. S. Bratan, H. Taan, and Y. L. Ismail, "Economics and Digital Business Review Meningkatkan Retensi Pelanggan Paket Data Internet Telkomsel di Kota Gorontalo melalui Strategi Customer Relationship Management (CRM)," *Econ. Digit. Bus. Rev.*, vol. 5, no. 1, pp. 47–56, 2024.

[11] N. Suryana, "Prediksi *Churn* Dan Segmentasi Pelanggan TV Berlangganan (Studi Kasus Transvision Jawa Barat)," *J. TEDC*, vol. 11, no. 2, pp. 185–191, 2019, [Online]. Available: https://ejournal.poltektedc.ac.id/index.php/tedc/article/view/77

[12] A. S. Wibowo, "Analisis *Churn* Nasabah Bank Dengan Pendekatan Machine Learning dan Pengelompokan Profil Nasabah dengan Pendekatan Clustering," *Konstr. Publ. Ilmu Tek. Perenc. Tata Ruang dan Tek. Sipil*, vol. 2, no. 1, pp. 30–41, 2024, doi: https://doi.org/10.61132/konstruksi.v2i1.43.

[13] A. Primajaya *et al.*, "*Random Forest* Algorithm for Prediction of Precipitation," *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, pp. 27–31, 2018.

[14] D. Marcelina, A. Kurnia, and T. Terttiaavini, "Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 293–301, 2023, doi: 10.57152/malcom.v3i2.952.

[15] KNIME Official, "KNIME Analytics *Platform*," 2023. https://www.knime.com/knime-analytics-*platform*

[16] J. Kurniawan *et al.*, *Analisa dan Visualisasi data*, 2023rd ed. Bandung: Widina Bhakti Persada Bandung, 2023.

[17] M. Ariandi *et al.*, "Analisis Visualisasi Data Kecamatan Kertapati menggunakan," *J. Jupiter,* vol. 14, no. 2, pp. 366–373, 2022.