INFORMATION
Volume 24, Number 3, pp.143-154

ISSN 1343-4500 eISSN 1344-8994

Consumer Behavior Analysis of Leathercraft Small and Medium-Sized Enterprises (SME) Using Market Basket Analysis and Clustering Algorithms

Dian Indri Purnamasari*, Asep Saepudin**, Vynska Amalia Permadi*** and Riza Prapascatama Agusdin***

*Department of Accounting, Faculty of Economy and Business,
Universitas Pembangunan Nasional Veteran Yogyakarta, Yogyakarta, Indonesia
E-mail: dian_indri@upnyk.ac.id

**Department of International Relations, Faculty of Social Science and Political Science,
Universitas Pembangunan Nasional Veteran Yogyakarta, Yogyakarta, Indonesia

*** Department of Informatics, Faculty of Industrial Engineering,
Universitas Pembangunan Nasional Veteran Yogyakarta, Yogyakarta, Indonesia

Abstract

Customers are essential to a business's existence so did effective customer management, which may increase companies' revenue. On the contrary, managing consumer preferences is not straightforward. Companies must prove their capacity to apply the most successful business strategies to match customer preferences. Thus, they will require a procedure for doing business research that is both compliant and capable of developing customer-driven resolutions. Consumer behavior analysis is a frequently used analytical technique in the business area to enhance marketing strategies or determine future corporate intentions. Consumer behavior can be investigated for various objectives, including customer transaction analysis and profiling. Consumer behavior analysis may be complicated for businesses that deal with enormous amounts of customer transactions history data. Thus, a data mining method might be required to aid in the inquiry. This study used the Apriori data mining technique to conduct a market basket analysis towards uncovering product transaction association rules. The K-Means clustering technique is then applied, as is the Elbow approach for identifying the ideal K to use in the customer segmentation clustering process. K-Means clustering has been able to find two dominant consumer groups through this research, whereas Apriori association identifies two product association rules with the highest possible confidence level.

Key Words: Consumer Behaviour Analysis, Consumer Segmentation, Market Basket Analysis

1. Introduction

Producing and identifying useful data is becoming essential for many businesses and industries interested in Customer Relationship Management (CRM) analysis in this digital era. One of the objectives of CRM is to gather as much information as possible about customers' preferences and behavior in order to provide the best possible service and maintain existing relationships. Thus, operating a CRM in concurrence with business research is beneficial throughout the customer life cycle. CRM is ideally defined as the cross-functional process of keeping an ongoing dialogue with the customer, leveraging all

interactions according to the customized treatment to increase customer retention and marketing activity effectiveness [1]. The CRM framework may be classified into two groups in terms of business planning: operational and analytical. The goal of operational CRM is to automate business operations.

In contrast, analytical CRM focuses on studying customers' descriptions, attitudes, interactions, and behaviors in order to assist company managers in comprehending consumer demands [2]. The analytics approach for customer relationship management (CRM) is commonly based on customer data. CRM's analytical capabilities have grown in prominence and are increasingly being considered by businesses. Analytical CRM is a data conversion cycle to create personalized marketing campaigns based on consumer behavior [2]. In business research, demographic, socioeconomic, consumer preference, or customer satisfaction are employed to maintain analytical CRM. Nowadays, analytical methodologies have evolved from a product-centric focus toward a customer-centric one, employing more conscientious CRM or data mining instruments. Data mining may be utilized across industries to assess customer behavior and purchase trends to manage CRM during a business cycle. For instance, by examining sales activity, notable trends or insights might be found. Consumer behavior analysis is necessary to decipher customer preferences in order to deliver the best service possible. For instance, by analyzing consumer data based on gender and age, it may be feasible to infer that persons of different ages have dissimilar preferences. Likewise, the customer's location might impact their choice of an item or color as well as the price range and style. Several types of research on customer satisfaction and behavioral analysis have been conducted, one of which makes use of data mining techniques such as association analysis [3] [4] [5]. Additionally, different methodologies or approaches have been discussed, including classification and clustering [6] [7], forecasting and regression [8] [9], and sequence discovery [10]—all of these researches have been evaluating actual data from the companies' databases.

As with prior research, this study will use data analysis to ascertain different CRM trends among leathercraft SMEs in Yogyakarta, Indonesia. Because they do not yet have connected databases, primary data from surveys will be used to develop CRM tools in the form of suggestions that may be applicable to their marketing plans for handcrafted products. This study aims to conduct a market basket analysis to deduce product transaction association rules using an Apriori data mining approach. Additionally, the K-Means clustering method is

used, as is the Elbow approach for determining the optimal K to utilize in the clustering process for customer segmentation.

2. Materials and Methods

2.1 Data sets

Primary data were collected through a consumer survey conducted by one of Yogyakarta's leather craft SMEs, questioning customers about their purchasing habits. Throughout the survey, 689 responses were collected. Prior to analyzing customer behavior by exploiting survey data, the following pre-processing steps are performed:

i. Transforming survey responses into the analytic format
ii. Data duplication elimination
iii. Developing neat and consistent survey responses that convey the same message but are worded differently

2.2 Data Mining Techniques

Data mining and machine learning are two of the most often used methodologies for data analysis. Data mining is critical in the KDD (Knowledge Discovery in Databases) process because it enables the discovery of patterns and the execution of analytical computations necessary to derive insight from data. CRISP-DM (Cross Industry Standard Method for Data Mining) is a data mining standard technique for completing an industry analysis that serves as the foundation for a business or research unit's problem-solving strategies [11]. The CRISP-DM [12] step-by-step is given as follows:

1. Comprehension of the business: Describes the business's objectives and requirements to transform into data mining issue specifications. Additionally, plans and strategies that will be developed to achieve these objectives are also defined.
2. Data comprehension: Begins with data collecting and continues with acquiring a thorough knowledge of the data, identifying data complexities, and identifying any intriguing pieces of data that could be leveraged to generate insight from the data.
3. Data preparation: Transforming raw data into the final data set (data that will be processed during the modeling step). This step may be repeated more than once, involving the selection of tables, records, and data attributes, as well as the cleansing and transformation of data in preparation for use as input in the modeling stage.
4. Modeling: Several modeling techniques will be chosen and implemented, and various parameters will be altered to obtain the expected results. Numerous strategies can be

utilized to solve the same data mining challenge. On the other hand, modeling techniques require a certain framework to ensure that reverting to the prior phase is still viable at this moment.

5. Evaluation: The model that previously has been developed is projected to perform properly. Hence, to confirm it, the model needs to be evaluated whether it meets the established objectives during the initial phase (data understanding). This step is an important phase to make sure that the business challenge is being investigated appropriately.
6. Development (implementation): Finally, the acquired knowledge or information will be structured and presented in such a way so that others can utilize it. It could be as easy as creating a simple report or as complex as iteratively developing data mining tools within the organization.

2.3 Market Basket Analysis and Association Rules

Market Basket Analysis aids businesses in developing profitable items. Han and Cheng [13] assert that market basket research can increase sales by focusing marketing efforts and optimizing shelf space. Market Basket Analysis's fundamental notion is the association of consumers' purchase decisions. For instance, when customers visit a supermarket, they are substantially more likely to purchase a basket of things, frequently spanning multiple product categories. By analyzing market basket information, individuals can analyze data on product categories and commodities that are likely to be purchased in conjunction and which products or product categories are distinctive. This understanding enables managers to create targeted measures that influence purchase behavior, such as growing public demand, promoting a specific product category, or advertising special deals to promote a specific product [14].

According to Larose [15], data mining can be divided into various subfields based on the activities or vocations performed; one of these is association. In data mining, the association is used to uncover characteristics that coexist. It is most typically referred to as shopping cart analysis in the business sector. Associations encompass the following illustrations in business and research:

1. Researching to determine the proportion of customers of cellular telecommunications carriers who are likely to respond positively to service enhancements.
2. Determine which things at the supermarket are frequently purchased concurrently and which are never.

2.4 Apriori Algorithm

The Apriori approach is the most widely used technique for finding patterns having a high likelihood of occurrence. In Apriori, the k-itemset candidates are formed using the (k-1)-itemset combination from the preceding candidate combination iteration. The Apriori algorithm can perform the pruning steps to the k-itemsets candidate whose subsets containing (k-1) items are not included in the high-frequency pattern by the length of k-1 [16].

The Apriori method is one of several algorithms that enables frequent-itemset searches through association rule schemes. The step in which association rules are extracted from data sources begins with discovering a set of frequent items or a group of items that repeatedly appear together. In the next step, we will define support and confidence, which are two significant association indicators. Support is the value or the percentage of occurrences of an item's combination. On the other hand, confidence is the value of certainty or the strength of the association rule's relationship between items. A brief explanation about support and confidence value and equations needed to perform the analysis is given below:

a) Analysis of high-occurrence patterns

The support value is calculated to identify a combination of objects, and the results are used to pinpoint pairs with a high occurrence level that match the minimum support criteria given. The following formula is used to determine an item's support value.

$$Support\ (A) = \frac{The\ amount\ of\ transactions\ consist\ of\ A}{The\ amount\ of\ Transactions} \quad (2.1)$$

While the following formula is used to get the support value for the two objects.

$$Support\ (A,B) = \frac{The\ amount\ of\ transactions\ consist\ of\ A\ and\ B}{The\ amount\ of\ Transactions} \quad (2.2)$$

b) Association Rules Formation Establishment

The following formula yields the confidence value for the "if A then B" rule, which means if A and B are on the same basket appears on a transaction.

$$Confidence\ (A|B) = \frac{The\ amount\ of\ transactions\ consist\ of\ A\ and\ B}{The\ amount\ of\ Transactions\ consits\ of\ A} \quad (2.3)$$

Upon discovering all high-occurrence patterns, association rules can be defined by selecting the itemset that satisfies the minimum confidence requirements.

2.5 Customer Segmentation and Clustering Algorithm

With a vast client database, data analytics can be used to determine the optimal business approach for sustaining strong customer connections and achieving CRM objectives by segmenting consumers. In business, customer segmentation is used to categorize clients based on shared factors such as age, gender, annual income, residence, or purchasing history. Customer segmentation studies can aid business owners in developing segment-specific business strategies.

Through the application of data mining, a clustering algorithm can be used to segment consumers. Clustering is a technique for categorizing vast volumes of data. Numerous clustering methods have been created in various fields of research, most notably the K-Means technique. K-Means is a method for segmenting an item into k distinct subregions. Each item must be initially assigned to a specific cluster and gradually shifted to a different cluster until placed at the correct cluster.

K-Means is a partitional algorithm, as it is dependent on selecting the initial number of K-groups and then defining the initials centroid value [17]. The K-Means approach generates a cluster database iteratively. It accepts an input of the desired number of beginning clusters and outputs the final number of clusters. There will be K initials and K endings if an algorithm is required to construct K clusters. At random, the K-Means approach will choose a pattern of K as the centroid's starting point. If the position of the new centroid does not change, the number of iterations required to reach the centroid cluster will be determined randomly by the first centroid cluster candidate. The value of K chosen as the initial centre will be determined using the Euclidean Distance formula, which determines the shortest distance between the centroid point and the data/object. Clustering occurs when data with the shortest or closest distance to the centroid form a cluster [18].

The elbow method is a technique for generating information on the optimal number of clusters by comparing the number of clusters that will form an elbow at a place to the percentage of clusters that will form an elbow [17]. This method generates ideas by selecting a cluster value and then adding it to a data model to determine the optimum cluster. Additionally, the percentage calculated from the calculation compares the number of clusters added using a graph as a source of information. The various percentages of each cluster value can be displayed. If the angle formed by the value of the first cluster and the value of the second cluster equals the angle formed by the value of the first cluster, or if the value has reduced the most, then the cluster value is the best [19].

3. Results and Discussion

3.1 Exploratory Data Analysis

The dataset derived from the leathercraft SME survey has 689 responses with the following data attributes:

a) Customer Name: This attribute is subsequently converted into customer IDs for analytical purposes, as we are not required to analyze the name in this research. As a result, we had 689 unique identifiers to represent various consumers.

b) Gender: Unfortunately, gender distribution in this survey is 100% dominated by all women customers. Nevertheless, we still include this item in the data analytic steps.

c) Birth year: Respondents of the survey's birth year vary from 2009 (12 years old) to 1958 or aged 63 approximately.

d) Occupations: This attribute describes our respondents' occupations, consist of 12 different professions. This attribute is transformed into 12 categorical values (0-11) for analytical purposes.

e) Monthly income: On the survey, we indicate five different monthly incomes categories. Hence, we transform the results into five categorical values (0-4) for analytical purposes.

f) The amount of owned products: We provide three different categories of choices to represent this particular question. Hence we transform the results into three categorical values (0-2) for analytical purposes.

g) Batik pattern: As a case study on this research, the handmade leather bag made by leathercraft SME offers more than 20 batik motifs that customers can order on the launched product. To be precise, we analyze 21 different batik patterns that customers have ordered. The patterns variance distribution is depicted in Figure 1. Parang pattern is the best seller pattern, followed by Truntum and Sidomukti patterns.

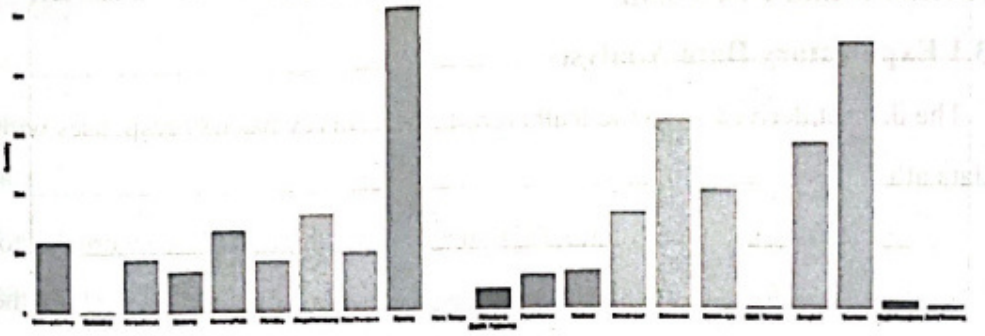

Figure 1. Quantities of the ordered Batik Patterns in leathercraft SME

h) Product color: This leathercraft SME deliver various color choices to fulfill customer preferences. The survey result (Figure 2) shows that the top three colors ordered are Havana, Borduk, and Black, which are all neutral color options.

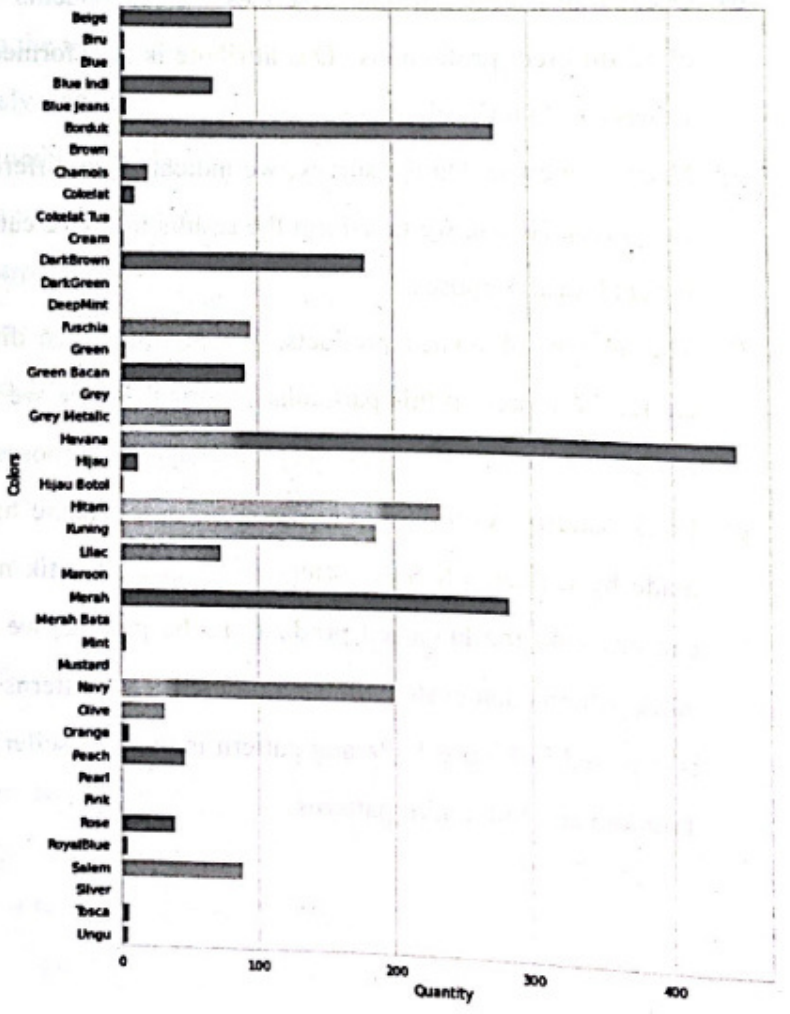

Figure 2. Quantities of the ordered product colors in leathercraft SME

i) Product material: Along with color and pattern variations, customers are offered to choose various leather materials. Berlea and Vitello are the most popular material alternatives among customers.

j) CRM-related questions: The survey sought to elicit information about customer transactions and included various questions about consumer satisfaction with SME products. Additionally, this study investigated these ten additional questions. Customer satisfaction is summarized using a Likert scale ranging from 1 to 5.

3.2 Apriori Association Rules

The Apriori method is implemented using Python to establish the association rules for batik patterns and colors based on market basket analysis. For batik motifs, the optimum rule yields minimal support at a value of 0.4 and a confidence level of 0.6. Moreover, the resulting rule is given in Table 1. The optimum rule created from the color preferences surveys data yields minimal support at a value of 0.06 and a confidence level of 0.01. Moreover, the resulting rule is given in Table 2.

Table 1. Association rules from batik patterns frequent itemset

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
1	(Truntum)	(Parang)	0.623188	0.730435	0.481159	0.772093	1.057032	0.025961	1.182786
0	(Parang)	(Truntum)	0.730435	0.623188	0.481159	0.658730	1.057032	0.025961	1.104146

Table 2. Association rules from colors frequent itemset

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
1	(Borduk)	(Havana)	0.398551	0.649275	0.3	0.752727	1.159334	0.041231	1.418372
0	(Havana)	(Borduk)	0.649275	0.398551	0.3	0.462054	1.159334	0.041231	1.118047

3.3 Customer Segmentation Clustering

Customer segmentation clustering has been performed by utilizing the Elbow method on the K-Means clustering algorithm. In this study, the Elbow technique is implemented through make use of Python tools. The Elbow technique has been applied to a dataset of 689 data points, and an iteration will begin with K = 1 and end with K = 9. The Elbow method's results indicate that the best value of K is only 2. Figure 3 below illustrates the SSE values for each iteration of the Elbow technique.

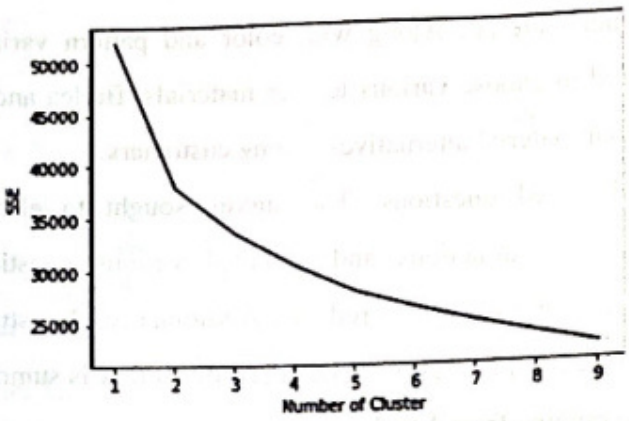

Figure 3. SSE values obtained from Elbow Method

To ensure that the number of clusters picked is correct, a silhouette score calculation is generated, the results are presented in Table 3. According to this score, the maximum number of clusters or ideal customer segmentation that may be advised matches the previous elbow method obtained, that is, two clusters.

Table 3. Obtained K-Means silhouette scores

	Silhouette Scores
2	0.155
3	0.095
4	0.101
5	0.080
6	0.080
7	0.083
8	0.076
9	0.054

4. Acknowledgments

The authors would like to thank the Institute for Research and Community Service at Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia, for providing funds for this research.

References

[1] J. F. Tanner Jr, M. Ahearne, T. W. Leigh, C. H. Mason and W. C. Moncrief, "CRM in Sales-Intensive Organizations: A Review and Future Directions, Journal of Personal Selling & Sales Management," *Journal of Personal Selling & Sales Management* , vol. 25, no. 2, pp. 169-180, 2005.

[2] E. Ngai, L. Xiu and D. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification," *Expert Systems with Applications,* vol. 36, no. 2, pp. 2592-2602, 2009.

[3] D. Nam, L. Junyeong and L. Heeseok , "Business analytics use in CRM: A nomological net from IT competence to CRM performance," *International Journal of Information Management,* vol. 45, pp. 233-245, 2019.

[4] B. Shim, C. Keunho and S. Yongmoo , "CRM strategies for a small-sized online shopping mall based on association rules and sequential patterns," *Expert Systems with Applications,* vol. 39, no. 9, pp. 7736-7742, 2012.

[5] J. U. Becker, G. Goetz and A. Sönke , "The impact of technological and organizational implementation of CRM on customer acquisition, maintenance, and retention," *International Journal of research in Marketing,* vol. 26, no. 3, pp. 207-215, 2009.

[6] S. M. S. Hosseini, M. Anahita and R. G. Mohammad , "Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty," *Expert Systems with Applications,* vol. 37, no. 7, pp. 5259-5264, 2010.

[7] V. Djurisic, L. Kascelan, S. Rogic and B. Melovic, "Bank CRM Optimization Using Predictive Classification Based on the Support Vector Machine Method," *Applied Artificial Intelligence,* vol. 34, no. 12, pp. 941-955, 2020.

[8] A. Ahani, Z. A. R. Nor and N. Mehrbakhsh , "Forecasting social CRM adoption in SMEs: A combined SEM-neural network method," *Computers in Human Behavior,* vol. 75, pp. 560-578, 2017.

[9] F. T. Bahari and M. S. Elayidom, "An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour," *Procedia Computer Science,* vol. 46, pp. 725-731, 2015.

[10] Y.-H. Hu, T. C.-K. Huang and Y.-H. Kao, "Knowledge discovery of weighted RFM sequential patterns from customer sequence databases," *Journal of Systems and Software,* vol. 86, no. 3, pp. 779-788, 2013.

[11] C. H. Cheng and Y. S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," *Expert Systems with Applications,* vol. 36, no. 3, pp. 4176-4184, 2009.

[12] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer and R. Wirth, "CRISP-DM 1.0 - Step-by-step data mining guide," CRISP-DM Consortium, 2000.

[13] J. Han, H. Cheng, D. Xin and X. Yan , "Frequent pattern mining: current status and future directions," *Data Mining and Knowledge Discovery,* vol. 15, p. 55–86, 2007.

[14] M. Madaio, R. Lasko, J. Cassell and A. Ogan, "Using Temporal Association Rule Mining to Predict Dyadic Rapport in Peer Tutoring,," *International Educational Data Mining Society,* pp. 318-323, 2017.

[15] D. T. Larose, Discovering Knowledge in Data: An Introduction to Data Mining, United States of America: John Wiley & Sons, Inc. , 2005.

[16] D. Solnet, Y. Boztug and S. Dolnicar, "An untapped gold mine? Exploring the potential of market basket analysis to grow hotel revenue," *International Journal of Hospitality Management* , vol. 56 , pp. 119-125, 2016.

[17] Madhulatha, T.S., "An Overview On Clustering Methods," *IOSR Journal of*

Engineering, vol. 4, pp.719-725, 2012.

[18] Agrawal, A., and Gupta, H., "Global K-Means (GKM) Clustering Algorithm: A Survey," *International Journal of Computer Applications*, vol. 59, no. 2, pp.20-24, 2013.

[19] Bholowalia, Purnima and Kumar, Arvind, "EBK-Means: A Clustering Techniques based on Elbow Method and K-Means in WSN," *International Journal of Computer Application,* vol. 9 no. 105, pp. 17-24, 2014.

*Corresponding author: Vynska Amalia Permadi.

Department of Informatics,

Universitas Pembangunan Nasional Veteran Yogyakarta,

Babarsari St. 2, Tambak Bayan, Depok, Sleman, Yogyakarta, Indonesia

E-mail: vynspermadi@upnyk.ac.id