

## ANALYSIS OF CUSTOMER SATISFACTION WITH THE APPLICATION OF DATA MINING USING THE K-MEANS CLUSTERING METHOD IN CV. GREEN PUBLISHER INDONESIA

**Dwi Febri Syawaludin, Muhammad Hatta, Kusnadi**

Faculty of Information Technology, Universitas Catur Insan Cendekia, Indonesia  
Email: febrisyawaludin445@gmail.com

### ABSTRACT

*In this digital era, data has become one of the most valuable assets for all companies. Data mining is a method that can be used to explore knowledge. The current research aims to analyze CV customer satisfaction. Green Publisher Indonesia by applying data mining techniques using the k-means clustering method. In this research, data will be collected regarding customer preferences and levels of satisfaction through sources such as online surveys. The data will be analyzed into the rapidminer 5.3 system using the k-means clustering algorithm. The research material used is customer data that has been obtained within a certain period by providing an Online Questionnaire (Google Form). In this research, researchers used one of the methods in Data Mining, namely the K-Means Clustering method. The results of the clustering execution using K-Means Clustering were 163 data resulting in 2 clusters with details of cluster 0 as 131 data or 80.36% and cluster 1 as 32 data or 19.64%. The results of the analysis show that there is a significant difference in the level of satisfaction between the two clusters.*

**KEYWORDS** *customer satisfaction, data mining, k-means clustering method*



*This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International*

### INTRODUCTION

CV. Green Publisher Indonesia is one of the publishing companies that has been contributing to the publishing industry in Indonesia for a long time. Along with the development of technology and increasingly fierce competition, maintaining customer satisfaction is very important for CV. Green Publisher Indonesia (Al Syahdan & Sindar, 2018). Customer satisfaction has become the main indicator of a company's success in maintaining and increasing its market share.

In this digital era, data is one of the most valuable assets for all companies.

**How to cite:**

**E-ISSN:**

Dwi Febri Syawaludin. et al (2025). Analysis of Customer Satisfaction with the Application of Data Mining Using the K-Means Clustering Method in CV. Green Publisher Indonesia. *Journal Eduvest*. 5(1), 963-970  
2775-3727

Data Mining is a method that can be used in exploring knowledge. By using Data Mining techniques, companies can understand more deeply about customer preferences and behavior, so that they can improve the services and products provided (Arifin et al., 2023).

Clustering is one of the techniques that can be used in Data Mining, which is a method of grouping data based on the similarity of its characteristics. By applying Data Mining clustering to a CV Green Publisher Indonesia customer data, the company can identify customer groups with similar characteristics (Kushardiyanti et al., 2023; Mardlotillah, 2020; Putra et al., 2020). This is able to support the company in designing more precise marketing strategies, improving customer satisfaction, and strengthening its position in the market.

This study aims to analyze CV customer satisfaction. Green Publisher Indonesia by applying Data Mining techniques using the K-Means clustering method. In this study, data will be collected regarding customer preferences and satisfaction levels through sources such as online surveys. The data will be analyzed into the rapidminer 5.3 system using the K-Means clustering algorithm using 2 clusters, namely cluster 0 and cluster 1 (Fau et al., 2017; Kody & Jollyta, 2022).

It is hoped that the results of this study can provide valuable insights for CV. Green Publisher Indonesia in improving service to customers so that customers can enter the criteria is very satisfied (Kusuma & Suwitho, 2015; Rohman, 2017; Santoso & Jamil, 2023). In addition, this research can make an important contribution to the field of Data Mining research in Indonesia, especially in the context of customer satisfaction analysis.

By understanding customer satisfaction more deeply through Data Mining clustering, CV. Green Publisher Indonesia can be better prepared to face increasingly fierce competition in the publishing industry. In addition, this research can also be an example for companies out there who want to utilize Data Mining as a tool to improve customer satisfaction and grow their business (Rizki et al., 2020; Siregar et al., 2017).

Based on the background above, the author wants to conduct research on CV. Green Publisher Indonesia entitled "Analysis of Customer Satisfaction with the Application of Data Mining Using the K-Means Clustering Method in CV Green Publisher Indonesia".

## **RESEARCH METHOD**

### **Data Collection Methods**

The following are the data collection techniques carried out by the author during this research, namely: (Sujarweni, 2014)

1. Literature Studies

This data collection technique is to search from various sources of information related to the research topic and the process that occurs in this research as a reference (Magdalena et al., 2023). In this case, the author collects various kinds of scientific journals, scientific articles, books, papers and other supporting documents related to the study of Customer Satisfaction Analysis with the Application of Data Mining Clustering in CV. Green Publisher Indonesia.

2. Observation

This data collection technique is by making direct observations of the objects in this study (Zaenurrohman, 2022). About the extent of Customer Satisfaction Analysis with the Implementation of Data Mining Clustering in CV. Green Publisher Indonesia. After this, all the information and data that are already owned obtained through this observation stage will be used as a guideline during the research process on the evaluation of risk management.

3. Interview

This data collection technique is by conducting interviews directly or by using liaison media. Such as by using a mobile phone or using online chat applications and using other electronic media (Rukajat, 2018). This interview was conducted by conducting a question and answer session by giving several questions to the resource persons related to the customer satisfaction survey manager in CV. Green Publisher Indonesia.

## RESULT AND DISCUSSION

### Program Test Results

The system was tested using the K-Means Clustering process using two methods, namely manual with Microsoft Excel and using the RapidMiner 5.3 application. The test results display the output of both methods, ensuring the consistency and validity of the process.

### RapidMiner 5.3 Test Results

The test results using the Rapidminer 5.3 application are as follows:

1. Centroid Table

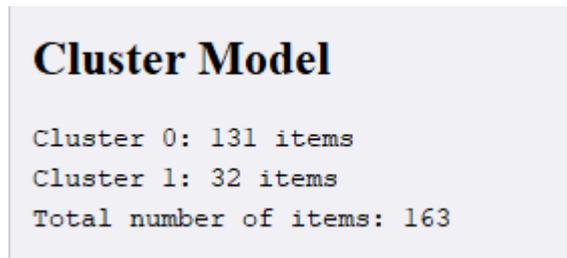
**Table 1. Centroid Table Display**

Attribute	cluster_0	cluster_1
Produk Publ	82.702	13.312
Kualitas/Mut	85.206	8.688
Pelayanan d	85.214	8.688
Harga produ	74.405	58.031

From the product attribute publish on time, the value for cluster 0 is 82,702 and cluster 1 is 13,312, the attribute quality/article quality value of cluster 0 is 85,206 and cluster 1 is 8,688, the attribute service and admin response value for cluster 0 is 85,214 and cluster 1 is 8,688, the product price attribute offered the value of cluster 0 is 74,405 and cluster 1 is 58,031.

2. Cluster Model

3.



**Figure 1. Model Cluster Display**

Based on the results of the RapidMiner 5.3 Aplikasi model cluster, namely cluster 0 with a total of 131 items and cluster 1 with a total of 32 items out of a total of 163 items.

**Microsoft Excel Test Results**

After passing the stages of the K-Means Clustering process using the Microsoft Excel application, the next process is to display the results of the application (Adani et al., 2019; Nabila et al., 2021).

1. Centroid Table

**Table 2. Centroid Table**

Attribute	Publish products on time	Quality/quality of articles	Admin service and response	Product prices offered
Cluster 0	82.7	85.2	85.2	74.4
Cluster 1	13.3	8.6	8.6	58.0

From the product attribute publish on time the value for cluster 0 is 82.7 and cluster 1 is 13.3, the attribute quality/quality article value of cluster 0 is 85.2 and cluster 1 is 8.6, the service attribute and admin response value for cluster 0 is 85.2 and cluster 1 is 8.6, the product price attribute offered the value of cluster 0 is 74.4 and cluster 1 is 58.0.

2. Cluster Model

**Table 3. Cluster Model**

Cluster	Sum
Cluster 0	131 data
Cluster 1	32 data
Total	163 data

Based on the results of the cluster model of the Microsoft Excel Application, namely cluster 0 with a total of 131 data and cluster 1 with a total of 32 data out of a total of 163 data.

**Discussion**

**Dataset**

The dataset is used as reference data to be processed and analyzed in this study. The customer dataset used totals 163. Some of the variables included in the dataset are as follows:

**Table 4. Variable Dataset**

No	Variable Name
1.	Publish products on time
2.	Quality/quality of articles
3.	Admin service and response
4.	Product prices offered

**Table 5. Database**

No	Publish products on time	Quality/quality of articles	Admin service and response	Product prices offered
1.	80	80	80	80
2.	80	82	80	83
3.	78	80	75	81
4.	100	100	100	100
5.	75	90	80	90
6.	78	88	82	90
7.	83	85	82	86
8.	75	70	75	70
9.	80	80	80	90
10.	81	81	81	85
.....				
163	90	85	88	81

### *Proses Clustering*

The algorithm used is K-Means Clustering while the clustering results are literacy carried out 6 times. The results of the clustering process will be processed and produce 2 previously formed clusters, namely cluster 0 for very satisfied and satisfied customers, cluster 1 for neutral or very dissatisfied and very dissatisfied customers. From these results, each field will have an average of the values of each cluster, so the average that goes into both clusters. For example, the price of the products offered has an average value of 74.4 for cluster 0 and a value of 58.0 for cluster 1. The details of the final value of the cluster centroids formed in each category field are shown in the following table:

**Tabel 6. Final Cluster Centroid**

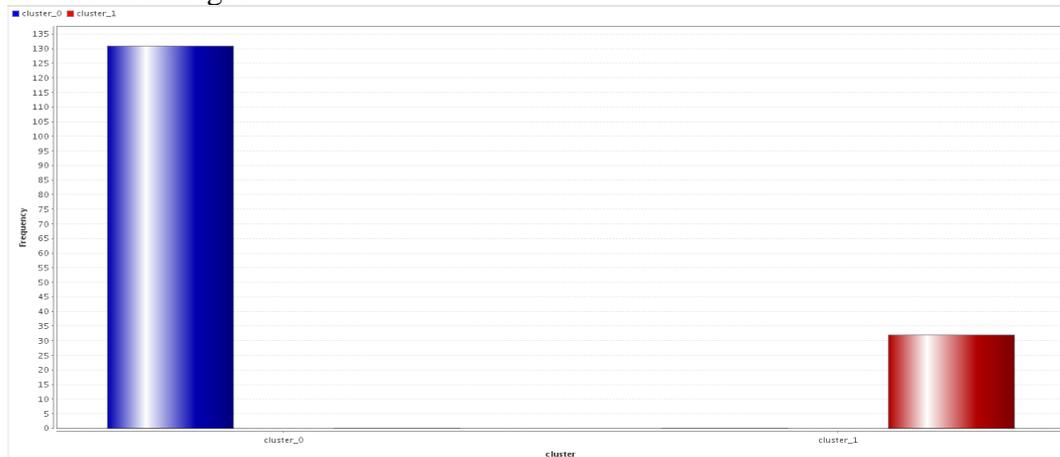
Attribute	Cluster	
	Cluster 0	Cluster 1
Publish products on time	82.702	13.312
Quality/quality of articles	85.206	8.688
Admin service and response	85.214	8.688
Product prices offered	74.405	58.031

The results of the clustering execution using K-Means Clustering as much as 163 data produce a cluster model shown in the following table:

**Table 7. Model Clusters and Percentages**

Cluster	Amount of Data	Percentage
Cluster 0	131	80,36%
Cluster 1	32	19,64%

After the cluster data is formed in Table 5.6 with each percentage, the percentage obtained from the results of 163 data is 100% divided into 2 clusters, namely cluster 0 with a total of 131 data and cluster 1 with a total of 32 data. Next, the data will be visualized in the form of diagrams or images to see the distribution of clusters. The results of the visualization of the distribution of data in each cluster are shown in Figure 5.3



**Figure 2. Cluster Distribution Visualization**

Cluster 0 is represented by blue with the largest number of data ranges of 131 while Cluster 1 is represented by red has the opposite data, namely 32 has less data.

## CONCLUSION

1. The application of the Data Mining method with the K-Means Clustering algorithm into the RapidMiner 5.3 application has been successfully implemented quickly and precisely using the rapidminer 5.3 application compared to the manual process using the Microsoft Excel application.
2. The results of the process from the Rapidminder 5.3 Application with 163 customer data were generated by cluster 2 with details of cluster 0 as many as 131 data or 80.36% and cluster 1 as many as 32 data or 19.64%.
3. The results of the analysis showed that there was a significant difference in the level of satisfaction between the two clusters. Cluster 0 shows a high level of satisfaction, based on the following attributes and criteria: timely publish products with criteria of 82.7 (very satisfied), quality/quality of articles with criteria of 85.2 (very satisfied), admin response service with criteria of 85.2 (very satisfied), product prices offered with criteria of 74.4 (quite satisfied). While cluster 1 shows a low level of satisfaction, based on

the following attributes and criteria: products published on time with criteria of 13.3 (very dissatisfied), article quality/quality with criteria of 8.6 (very dissatisfied), admin response service with criteria of 8.6 (very dissatisfied), product prices offered with criteria of 58.0 (quite satisfied).

## REFERENCES

- Adani, N. F., Boy, A. F., & Syahputra, R. (2019). Implementasi Data Mining Untuk Pengelompokan Data Penjualan Berdasarkan Pola Pembelian Menggunakan Algoritma K-Means Clustering Pada Toko Syihan. *Jurnal Cyber Tech*, 2(5).
- Al Syahdan, S., & Sindar, A. (2018). Data mining penjualan produk dengan metode apriori pada Indomaret Galang Kota. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 1(2).
- Arifin, Z., Ariantini, M. S., Sudipa, I. G. I., Chaniago, R., Dwipayana, A. D., Adhicandra, I., Ariana, A. A. G. B., Yulianti, M. L., Rumata, N. A., & Alfiah, T. (2023). *GREEN TECHNOLOGY: Penerapan Teknologi Ramah Lingkungan Berbagai Bidang*. PT. Sonpedia Publishing Indonesia.
- Fau, A., Mesran, M., & Ginting, G. L. (2017). Analisa Perbandingan Boyer Moore Dan Knuth Morris Pratt Dalam Pencarian Judul Buku Menerapkan Metode Perbandingan Eksponensial. *Jurnal TIMES*, 6(1), 12–22.
- Kody, J., & Jollyta, D. (2022). Aplikasi Clustering Tingkat Kepuasan Pelanggan untuk Mengukur Kepuasan Pelanggan: Uji Black Box, Reliabilitas dan Validitas. *SEMINAR NASIONAL INFORMATIKA (SENATIKA)*, 6(3), 1–12.
- Kushardiyanti, D., Khotimah, N. K., & Kom, S. I. (2023). *Praktik & Strategi Manajemen Industri Media Massa (Studi Lapangan Manajemen Industri Media Massa Lokal Wilayah Cirebon)*. CV Jejak (Jejak Publisher).
- Kusuma, R. W., & Suwitho, S. (2015). Pengaruh Kualitas produk, harga, fasilitas dan emosional terhadap kepuasan pelanggan. *Jurnal Ilmu Dan Riset Manajemen (JIRM)*, 4(12).
- Magdalena, L., Mar'atus Solikhah, A. M., Komarudin, A. M., Renaldy, R., Md, A., Fahrezi, R. B., Md, A., Sukma Hendrian, A. M., Karsa, A. H. A. N., & Md, A. (2023). *AGILE PRACTICE: PERAN TRELLO DALAM ADOPTSI AGILE SCRUM PADA PENGEMBANGAN IDE BISNIS START UP*. CV. Green Publisher Indonesia.
- Mardlotillah, N. I. (2020). Manajemen risiko keselamatan dan kesehatan kerja area confined space. *HIGEIA (Journal of Public Health Research and Development)*, 4(Special 1), 315–327.
- Nabila, Z., Isnain, A. R., Permata, P., & Abidin, Z. (2021). Analisis data mining untuk clustering kasus covid-19 di Provinsi Lampung dengan algoritma k-means. *Jurnal Teknologi Dan Sistem Informasi*, 2(2), 100–108.
- Putria, H., Maula, L. H., & Uswatun, D. A. (2020). Analisis proses pembelajaran dalam jaringan (daring) masa pandemi covid-19 pada guru sekolah dasar. *Jurnal Basicedu*, 4(4), 861–870.
- Rizki, M., Devrika, D., & Umam, I. H. (2020). Aplikasi Data Mining dalam penentuan layout swalayan dengan menggunakan metode MBA. *Jurnal Teknik Industri: Jurnal Hasil Penelitian Dan Karya Ilmiah Dalam Bidang Teknik Industri*, 5(2), 130–138.

- Rohman, A. (2017). *Pengaruh Kualitas Pelayanan Terhadap Kepuasan Pelanggan Bengkel Mr Montir Citayam*.
- Rukajat, A. (2018). *Pendekatan penelitian kuantitatif: quantitative research approach*.
- Santoso, D. B., & Jamil, H. (2023). *Ekonomi Industri: Tinjauan Teori dan Kebijakan*. MNC Publishing.
- Siregar, A. M., Kom, S., Puspabhuana, M. K. D. A. N. A., Kom, S., & Kom, M. (2017). *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV Kekata Group.
- Sujarweni, W. (2014). *Metodologi penelitian: Lengkap, praktis, dan mudah dipahami*.
- Zaenurrohman, I. J. A. (2022). KEKUATAN DAN KELEMAHAN RISET BERBASIS DIGITAL. *Metode Riset Berbasis Digital: Penelitian Pasca Pandemi*, 47.