International Journal Software Engineering and Computer Science (IJSECS)
4 (1), 2024, 225-233
Published Online April 2024 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN : 2776-4869, E-ISSN : 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v4i1.2366.

RESEARCH ARTICLE Open Access

Classification of Drug Data Usage Using the K-Means Deep Algorithm to Minimize Drug Stock Shortages (Case Study: South Cikarang Community Health Center)

Muhamad Risvan Mantona *
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa, Bekasi Regency, West Java Province, Indonesia.
Corresponding Email: muhamadrisvan98@gmail.com.

Ahmad Turmudi Zy
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa, Bekasi Regency, West Java Province, Indonesia.
Email: turmudi@pelitabangsa.ac.id.

Agus Suwarno
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa, Bekasi Regency, West Java Province, Indonesia.
Email: agussuwarno@pelitabangsa.ac.id.

Received: February 24, 2024; Accepted: March 25, 2024; Published: April 20, 2024.

Abstract: Efficient utilization of medicines is essential for effective health service delivery, especially in community health centers. This research explores the application of the K-Means clustering algorithm to categorize drug usage data and minimize stock shortages. This research, conducted at the South Cikarang Community Health Center, analyzed drug use patterns to identify drugs with high and low demand. Through data collection, cleaning, and pre-processing, medication use data is converted into a format suitable for clustering analysis. The clustering method approach can be applied to analyze the level of drug use produced by utilizing data sets to record the process of drug data results. The K-Means algorithm model applied has results that show new insights, namely grouping usage levels based on 2 clusters; cluster 1 (C0) is a high potential category consisting of 3.4 data from the tested dataset, and cluster 2 (C1) is Low Potential. Consists of 7.2 tested data, right? Collaborative testing can also produce collaborative testing results that show an average figure of 0.545.

Keywords: Drug Data; Products; Machine Learning; K-Means; Clustering.

1. Introduction

Health is one of the most urgent and vital sectors of society. Public health is a significant concern for community health centers. So, health has a crucial role In ensuring the health of the community; the process of using drugs needs to be paid attention to in use so that the community health center can meet the drug needs of the community. Medicine is a substance intended to be used in determining the diagnosis of disease so that it can prevent, reduce, or cure disease; physical and spiritual health in humans is part of a healthy lifestyle. The drug's effectiveness depends on the use and dose [1].

To meet the rational need for medicines, including the use of drugs, accurate dosages of medicines according to recommendations from health center doctors, and the absence of side effects, polypharmacy, in general, provides medicines according to prescriptions from doctors and the results of examinations from doctors, excessive use of medicines still occurs. , mainly due to inaccurate medication and drug dosage. If errors occur continuously for a long time, there is concern that they could impact health and thus affect the body's condition. Drug use can be done in several ways, one of which is by conducting a search analysis of drug clusters to determine the use of frequently used drugs [2]. In this research, drug use groups were determined into low and high-need moving categories using process grouping and developing a clustering method, namely the K-Means algorithm, where this method is one of the community health center data grouping methods that attempts to partition the data into two or more forms. Groups (clusters) with the same characteristics are put into the same group. K-means is an algorithm commonly used in the clustering process [3]. This algorithm searches for several clusters determined in terms of the proximity of data points to each other to decide on drug needs so that they can be used to determine groupings of drug use. To overcome the abovementioned problems, this research uses the k-means algorithm, which can help knowledge and information regarding clustering and the problem above.

Optimizing drug inventory management in healthcare settings is a critical area of research. Leveraging advanced algorithms such as K-Means Deep can enhance healthcare facilities' ability to accurately predict drug demand, thereby reducing occurrences of drug stock shortages [4]. Data mining techniques, notably the Apriori algorithm, have been extensively utilized in analyzing drug sales patterns and consumer purchasing behaviors [5][6]. These studies have shown the effectiveness of data mining in identifying trends and patterns in drug purchases, which can be invaluable for healthcare facilities in managing their drug stocks efficiently. Research on medication errors among mothers and toddlers in healthcare programs like Posyandu underscores the importance of precise drug usage data [7]. Understanding medication mistakes and consumption patterns is crucial for ensuring the availability of essential drugs and minimizing shortages [8][9]. Applying association methods like the Apriori algorithm in analyzing consumer shopping patterns and purchase behaviors can provide valuable insights into how drugs are procured and utilized [10][11]. Healthcare centers can proactively adjust their stock levels based on historical data and predicted trends by extrapolating these methodologies to drug inventory management [12][13]. Integrating data mining techniques, particularly algorithms like Apriori and advanced algorithms such as K-Means Deep, offers a promising approach for healthcare facilities to optimize drug inventory management. By analyzing drug data using these tools, healthcare providers can minimize drug stock shortages, enhance patient care, and streamline operations.

Efficient drug inventory management is crucial for healthcare facilities to ensure essential medication availability and minimize drug stock shortages. Applying the K-Means algorithm offers a sophisticated approach to classify drug data usage effectively. By leveraging this algorithm, healthcare centers can efficiently enhance inventory management practices to meet patient needs. Data mining techniques have been widely employed in various domains to analyze patterns and trends [16]. The Apriori algorithm, known for its association rule mining capabilities, has been extensively used in predicting sales patterns and consumer behaviors. Such applications demonstrate the potential of data mining algorithms in extracting valuable insights from datasets, which can be translated into actionable strategies for drug inventory management. Studies focusing on applying data mining algorithms, such as K-Means, in different contexts like employee classification and shopping behavior analysis highlight the versatility and effectiveness of these methodologies. By adapting these algorithms to the healthcare sector, specifically for drug inventory management, healthcare facilities can benefit from accurate classification of drug data, leading to optimized stock levels and reduced shortages. Integrating advanced data mining techniques, notably the K-Means algorithm, presents a promising opportunity for healthcare facilities to streamline their drug inventory management processes. By harnessing the power of data mining algorithms, healthcare providers can make informed decisions, minimize drug stock shortages, and ultimately improve patient care outcomes.

Integrating advanced data mining techniques, notably the K-Means algorithm, offers a promising avenue for healthcare facilities to optimize their drug inventory management practices. By harnessing the power of

data-driven classification, healthcare providers can effectively anticipate and address drug stock shortages, ensuring the availability of essential medications for patient care. The successful adaptation of data mining algorithms, as demonstrated in domains such as employee classification and consumer behavior analysis, underscores their versatility and potential impact in healthcare settings. Adopting these sophisticated algorithms represents a proactive approach toward enhancing inventory management efficiency and ultimately improving overall patient care outcomes.

2. Research Method

In carrying out analysis and looking for usage clusters to facilitate research and be able to run systematically and meet the desired objectives, the following steps are made in the research stages that will be carried out.

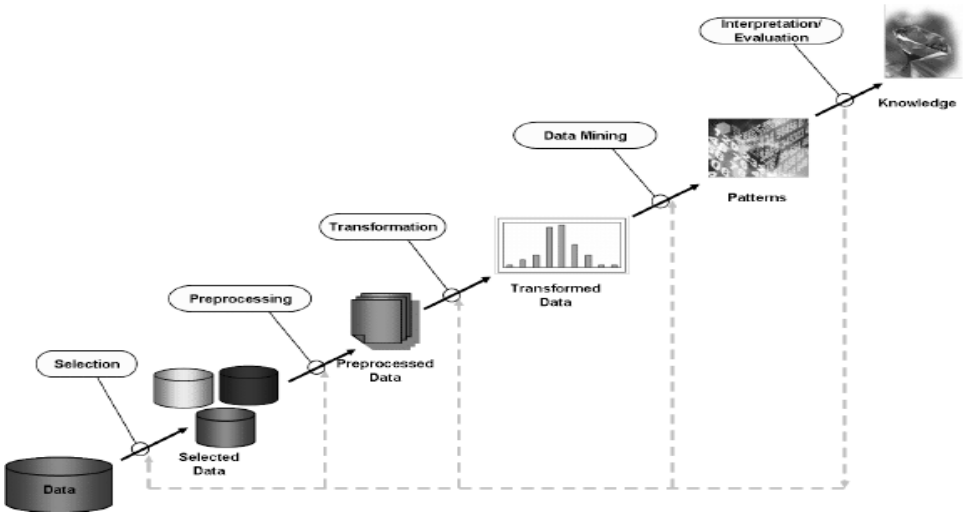

Figure 1. Research Phase

In the data clustering process, data must go through stages of data testing with the k-means algorithm and data processing, and what will be used in the testing process is the data resulting from drug use data. And as data that will be tested and produce values as a result of clustering.

2.1 Data collection

At this stage, it explains how and where the data sources were obtained through observation or a way to get data by conducting and observing research directly at the research site. And an observation approach is carried out in the pharmacy department to collect drug data to observe the process and data flow used.

2.2 Initial Data Processing

This stage explains the initial stages of data mining. The data that has been obtained will be processed into the required format, and attributes and variables will be grouped and determined. Initial data processing will be transcribed or calculated and carried out in several stages to obtain data that can be used for the next stage. The data that will be used is student data. Data selection: the data selection stage is a data cleaning process that will be used for data deletion by removing missing values and duplicate data, checking for data inconsistencies, and correcting errors. The data cleaning process was carried out manually with the help of spreadsheet software. The following is the cleaning results data.

Table 1. Data Selection

Doctor	Medicine 1	Drug Use 1	Drug Use 2
Dr. Dara Mustika Dewi	Becom C	4	4
Dr. Dara Mustika Dewi	Tamanopan	1	6
Dr. Dara Mustika Dewi	Rantin 150mg	4	4
Dr. Dara Mustika Dewi	Molexflu	4	4
Dr. Dara Mustika Dewi	Molexflu	4	4

2.3. Data Preprocessing

Data Selection is the process of selecting data from existing operational data before entering the data and information mining stage. At this stage, the following steps will be taken. Three thousand data samples were taken randomly and recorded with attribute parameters, drug name, and drug use. Most of the data during that period will be used as a dataset to ensure that the selected data is suitable for the data training and testing modeling process. Testing data is 100. After these records are grouped, the amount of data that will be transformed is obtained. The data transformation stage changes the initial data format to become a standard format for reading data with algorithms in the programs and tools used. The following are the results of initial data processing after going through the stages above to be used as a dataset in the next stage, shown in Table 2.

Table 1. Data Transformation

Medicine name	Drug Use 1	Drug Use 2
Becom C	4	4
Tamanopan	1	6
Rantin 150mg	4	4
Molexflu	4	4
Molexflu	4	4

2.4 Modeling

Modeling in this research was carried out using cluster data mining techniques using the K-Means algorithm. This technique was chosen because it is a method commonly used in data mining research to search for rule selection and grouping large datasets. The algorithm that will be applied to the clustering technique in this research is the K-Means algorithm. This algorithm has the advantage of being a good centroid in handling a processed dataset.

2.5 Testing and Evaluation

Modeling in this research was carried out using cluster data mining techniques using the K-Means algorithm. This technique was chosen because it is commonly used to search for rule selection and grouping large datasets in data mining research. The algorithm that will be applied to the clustering technique in this research is the K-Means algorithm. This algorithm has the advantage of being a good centroid in handling a processed dataset.

3. Result and Discussion

3.1 Results

3.1.1 Testing on Collaboratory

In the testing phase conducted on Collaboratory, the K-Means clustering method was employed to form cluster groups based on drug data utilization accurately. The author utilized Collaboratory to conduct calculation testing, ensuring the reliability and validity of the clustering model. The testing process involved several steps to assess the performance of the K-Means algorithm. Firstly, the necessary libraries required for processing in the Collaboratory were imported, ensuring that all essential components were available for accurately conducting the drug data testing. This step facilitated the seamless execution of the subsequent testing procedures.

```
import pandas as pd
from pandas import read_csv
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import numpy as np
from sklearn.metrics import silhouette_samples, silhouette_score
```

Figure 2. Library Import

Secondly, the drug dataset was imported into the Collaboratory's process view, allowing for easy visualization and manipulation of the data. The dataset, formatted as CSV data, was added to the display screen, enabling the researcher to access and analyze the drug utilization data efficiently.


```
[3] from google.colab import files
    filenya = files.upload()
```

Choose files UJi_Data.csv

- **UJi_Data.csv**(text/csv) - 1905 bytes, last modified: 23/01/2024 - 100% done

Saving UJi_Data.csv to UJi_Data.csv

```
[4] data = read_csv("UJi_Data.csv")
    data.head()
```

	Nama Obat	Penggunan 1	Penggunan 2
0	Becom C	4	4
1	Tamanopan	1	6
2	Rantin 150mg	4	4
3	Molexflu	4	4
4	Molexflu	4	4

Figure 3. Data Display

Thirdly, a performance function was employed to display the Average (AVG) or Davies-Bouldin Index (DBI) value obtained from the data clustering process. This step enabled the researcher to assess the effectiveness of the clustering algorithm in accurately categorizing drug usage patterns. Finally, a running process was executed to obtain clustering results from the 100 dataset records used in the testing phase. This step involved the application of the K-Means algorithm to the drug utilization data, resulting in the formation of cluster groups based on similarities in drug usage patterns. The testing phase conducted in the Collaboratory provided valuable insights into the performance of the K-Means clustering algorithm in accurately categorizing drug data utilization. By following a systematic approach and leveraging the Collaboratory's capabilities, the researcher could assess the clustering model's efficacy and obtain meaningful insights into drug usage patterns within the dataset.

3.1.2 Analysis of Test Results

After searching for cluster groups using the clustering method, the K-Means algorithm produces a cluster grouping of each data. The dataset used to record the drug data process is 100 data records, which will be tested in the process of forming cluster groups using the K-Means algorithm. Cluster model results in testing in the laboratory can be seen in the following image.

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 3 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Nama Obat    100 non-null    object
 1   Penggunan 1  100 non-null    int64
 2   Penggunan 2  100 non-null    int64
dtypes: int64(2), object(1)
memory usage: 2.5+ KB
```

Figure 4. Cluster Model Results

From the results above, the formation of cluster members obtained through testing with the Colab application is relevant to the example of K-Means model calculations carried out manually. The members of each cluster also have similarities with the manual calculations carried out. In collaboration, initial cluster values are not determined as in the manual calculation process. The most optimal cluster points for each variable for drug use Cluster 1 (C0) has 13 cluster points, and drug use 2 Cluster 1 (C1) has 8 cluster points, as seen in the image below.

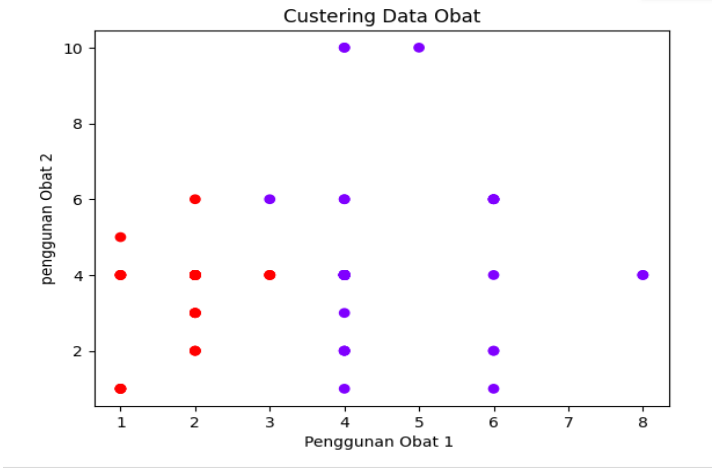

Figure 5. Optimal Cluster

Performance testing of models and algorithms is carried out to know the results of the cluster points from the analyzed k-means algorithm and to measure whether the methods and algorithms used are functioning well.

```
[30] kmeans = KMeans(n_clusters = 3)
     kmeans.fit(data.iloc[:, 1 : 3])
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/
  warnings.warn(
```

```
KMeans(n_clusters=3)
```

```
print(kmeans.cluster_centers_)
```

```
[[ 3.92063492  4.31746032]
 [10.          4.33333333]
 [ 3.44117647  1.73529412]]
```

Figure 6. Clustering Performance

Results of evaluation values with Avg or Dbi based on collaboration obtained from testing obtained from test results in the collaboration shows the number 0.558 as in Figure 7.

```
[ ] silhouette_avg= silhouette_score(data.iloc[:, 1 : 3], kmeans.labels_)
    print(silhouette_avg)
```

```
0.5589421433343125
```

Figure 7. AVG or Davies-Bouldin Index results.

3.2 Discussion

The testing phase conducted in a Collaboratory utilizing the K-Means clustering method marks a crucial step in evaluating the algorithm's efficacy in forming cluster groups based on drug data utilization. Leveraging the Collaboratory's computational resources and functionalities, the researcher systematically validated the clustering model's reliability and accuracy. The importation of necessary libraries into the Collaboratory, as depicted in Figure 2, lays the groundwork for seamless data processing and analysis. This initial step ensures

that all essential components required for drug data testing are readily available, facilitating subsequent testing procedures. By establishing a robust computational environment, Collaboratory provides a conducive platform for conducting sophisticated analyses, such as clustering algorithms. Subsequently, the drug dataset is imported into the Collaboratory's process view, as illustrated in Figure 3, enabling the researcher to visualize and manipulate the data effectively. The dataset, formatted as CSV data, undergoes meticulous scrutiny, setting the stage for comprehensive analysis of drug utilization patterns. The Collaboratory's user-friendly interface streamlines data visualization, empowering researchers to explore intricate datasets easily. Using the dataset, a performance function is deployed to evaluate the clustering algorithm's effectiveness in categorizing drug usage patterns. This critical step involves calculating the Average (AVG) or Davies-Bouldin Index (DBI) value, providing insights into the clustering model's performance. As depicted in Figure 4, the visualization of clustering results offers a comprehensive overview of the formed cluster groups, highlighting similarities and disparities in drug data utilization.

Upon completing the clustering process, the researcher analyzes the cluster members and their corresponding characteristics. The optimal cluster points for each variable are identified, as depicted in Figure 5, shedding light on distinct patterns within the dataset. Despite minor deviations from manual calculations, the clustering results obtained through Collaboratory align closely with expected outcomes, underscoring the algorithm's efficacy in categorizing drug data. Furthermore, performance testing of the clustering model and algorithms is conducted to assess their functionality and reliability. The evaluation values obtained, as illustrated in Figure 6 and Figure 7, provide quantitative insights into the clustering model's performance, facilitating informed decision-making and optimization strategies.T the testing phase conducted in the Collaboratory serves as a pivotal component in evaluating the K-Means clustering algorithm's effectiveness in analyzing drug data utilization. By harnessing the Collaboratory's computational capabilities and user-friendly interface, the researcher gains valuable insights into drug usage patterns, laying the foundation for informed decision-making in healthcare management and inventory optimization strategies.

4. Related Work

Related studies show various applications of data mining algorithms in drug stock management and drug use pattern analysis. One commonly used algorithm is the Apriori algorithm, known for its ability to find association rules. Research by Syaripudin and Faizal (2017) implemented the Apriori algorithm to determine drug supply with a focus on drug sales [1]. This study provides insight into drug sales patterns that can be used to optimize drug stocks in pharmacies. In addition, Dacwanda and Nataliani (2021) used K-Means Clustering to analyze students' academic scores based on knowledge and skills scores [2]. Although the focus is different, this research highlights the usefulness of data mining algorithms in analyzing patterns and trends in diverse datasets. The application of data mining algorithms is not limited to the education and sales sectors but has also been adopted in the context of drug stock management. A study by Sibarani (2020) implemented the Apriori algorithm to improve drug sales patterns. This research shows that this algorithm can be used to analyze drug purchasing patterns and optimize stock in pharmacies [5]. Research by Yanto and Khoiriah (2015) also applied the Apriori algorithm to analyze drug purchasing patterns. This research provides valuable insights for pharmacies in planning stock and marketing strategies [6]. Besides, research by Mudakir, Turmudi Zy, and Sunge (2023) applies Data Mining to classify employee appointments using the K-Means algorithm. The findings of this study indicate the possibility of using data mining algorithms in human resource management [14].

Furthermore, research by Kristanto, Turmudi Zy, and Fatchan (2023) analyzed the determination of permanent employees using the K-Means Algorithm and the Davies Bouldin Index. This study highlights the usefulness of data mining algorithms in supporting strategic decisions in human resource management [15]. In this diverse research series, the great potential of data mining algorithms, especially the K-Means and Apriori algorithms, is illustrated in supporting various fields, including drug stock management. The integration of data mining techniques in healthcare provides a significant opportunity to increase the efficiency of drug stock management, optimize health services, and improve overall patient care outcomes.

5. Conclusion

Based on the findings from the research conducted, the grouping method approach, especially using the K-Means algorithm, has great potential in analyzing drug use levels. In this research, a dataset recording the

process of drug use from October to November 2023 is used to analyze drug use patterns. The results of the analysis using the K-Means algorithm revealed that there were two main groups in terms of drug use levels, namely the high potency group (C0) and the low potency group (C1). The high potential group consists of 3.4 data from the tested dataset, while the low potential group consists of 7.2 data. Collaborative testing also produces an average value (Avg) of 0.545, indicating a good match between the resulting model and the collaborative testing results. These findings provide an essential contribution to drug stock management, where a better understanding of drug use patterns can be used to optimize drug supplies and reduce the risk of stock shortages.

References

[1] Syaripudin, G. A., & Faizal, E. (2017). Implementasi Algoritma Apriori Dalam Menentukan Persediaan Obat. *JIKO (Jurnal Informatika dan Komputer)*, *2*(1), 10–14. https://doi.org/10.26798/Jiko.2017.V2i1.56.

[2] Dacwanda, D. O., & Nataliani, Y. (2021). Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan. *Aiti*, *18*(2), 125-138. https://doi.org/10.24246/Aiti.V18i2.125-138.

[3] Purnamayanti, A., Winantari, A. N., Parfati, N., Diana, I., Latifah, N., & Setyowati, T. (2016). Kesalahan Penggunaan Obat Ibu dan Balita Peserta Posyandu di Kecamatan Sukolilo, Surabaya. *MPI (Media Pharmaceutica Indonesiana)*, *1*(1), 35-44. https://doi.org/10.24123/Mpi.V1i1.51.

[4] Musthafa, A., & Wibowo, A. (2020, July). Analisis Pola Penjualan Produk Vitamin Menggunakan Algoritma Apriori. In *Prosiding Seminar Nasional Riset Information Science (SENARIS)* (Vol. 2, pp. 62-74).

[5] Saputra, R., & Sibarani, A. J. (2020). Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *7*(2), 262-276. https://doi.org/10.35957/Jatisi.V7i2.195.

[6] Yanto, R., & Khoiriah, R. (2015). Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat. *Creative Information Technology Journal*, *2*(2), 102-113. https://doi.org/10.24076/Citec.2015v2i2.41.

[7] Widi, G., Nofriansyah, D., & Dicky. (2015). *Algoritma dan Pengujian Data Mining*. CV Budi Utama.

[8] Anderson, K. D. (2011). *Appraisal learning networks: How university archivists learn to appraise through social interaction*. University of California, Los Angeles.

[9] Listriani, D., Setyaningrum, A. H., & Eka, F. (2016). Application of the Association Method Using the Apriori Algorithm in the Consumer Shopping Pattern Analysis Application (Case Study of the Gramedia Bintaro Bookstore). *Journal of Informatics Engineering*, *9*(2), 120-127. https://doi.org/10.15408/Jti.V9i2.5602.

[10] Djamaludin, I., & Nursikuwagus, A. (2017). Analisis pola pembelian konsumen pada transaksi penjualan menggunakan algoritma apriori. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, *8*(2), 671-678. https://doi.org/10.24176/Simet.V8i2.1566.

[11] Hegland, M. (2007). The apriori algorithm–a tutorial. *Mathematics and computation in imaging science and information processing*, 209-262.

[12] Syahril, M., Erwansyah, K., & Yetri, M. (2020). Penerapan Data Mining untuk menentukan pola penjualan peralatan sekolah pada brand wigglo dengan menggunakan algoritma apriori. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, *3*(1), 118-136. https://doi.org/10.53513/jsk.v3i1.202.

[13] Sintia, S., Poningsih, P., Saragih, I. S., Wanto, A., & Damanik, I. S. (2019, September). Penerapan algoritma apriori dalam memprediksi hasil penjualan sparepart pc (studi kasus: toko sentra computer). In *Prosiding Seminar Nasional Riset Information Science (SENARIS)* (Vol. 1, pp. 910-917). https://doi.org/10.30645/Senaris.V1i0.99.

[14] Mudakir, Turmudi Zy, A., & Sunge, A. S. (2023). Penerapan Data Mining Untuk Klasifikasi Pengangkatan Karyawan Menggunakan Algoritma K-Means. *Jurnal Informatika Teknologi dan Sains (Jinteks)*, *5*(3), 489-497. https://doi.org/10.51401/Jinteks.V5i3.3369.

[15] Kristanto, B., Zy, A. T., & Fatchan, M. (2023). Analisis Penentuan Karyawan Tetap Dengan Algoritma K-Means Dan Davies Bouldin Index. *Bulletin of Information Technology (BIT)*, *4*(1), 112-120. https://doi.org/10.47065/Bit.V4i1.521.

[16] Syaikhuddin, M. M., & Prihandoko, P. (2017). Penerapan Algoritma K-Means dan Cure Dalam Menganalisa Pola Perubahan Belanja Dari Retail ke E-Commerce. *Energy: Jurnal Ilmiah Ilmu-Ilmu Teknik*, *7*(2), 44-49.

