IMPLEMENTATION OF DATA MINING TO CATEGORIZE DAILY SALES TRANSACTION DATA USING K-MEANS CLUSTERING

Firahmi Rizky[1], Yohanni Syahra[2], Hevlie Winda Nazry S[3], Jufri Halim[4], Niver Niawati Lo'i[5]

[1,2,3]Universitas Muhammadiyah Sumatera Utara
[4,5]STMIK Trigurna Dharma
[*1]*email: firahmirizky@umsu.ac.id*

Abstract: This research aims to implement Data Mining techniques using the K-Means Clustering method to group daily sales transaction data at Indomaret. This clustering aims to determine Indomaret's sales profit every day.
The data used in this research is daily sales transaction data at Indomaret. The K-Means Clustering process begins with random selection of cluster center points using the K-Means method and grouping data based on relevant attributes.
The results of this study show that K-Means Clustering can group daily sales transaction data effectively, producing several clusters that have unique characteristics. Each cluster represents a different market segment, which enables Indomaret management. Thus, the implementation of data mining using k-means clustering is proven.

Keywords*: K-Means Clustering, Daily Sales Transaction, Indomaret, Data Mining clustering*

Introduction

The development of information and communication technology has had a significant impact on various sectors, including the trade sector. Today's retail companies, such as Indomaret, not only face large amounts of sales data but also need to understand patterns and trends from the data. To gain deeper insights, companies need to implement sophisticated data analysis techniques. One of the techniques that can be used is data mining.

Indomaret which is located at Jalan Jamin Ginting k.m 8.5 which was established in 2005. There is what the problem is.

Data mining is the mining or discovery of new information by looking for certain patterns or rules from a very large amount of data [1]. By applying data mining techniques, companies can optimize sales transaction reports at Indomaret. In data clustering techniques such as K-Means Clustering, it can be used to group daily sales transaction data based on the K-Means method.

Literature Review

Data Mining

Data mining is the process of finding interesting patterns or information on selected data using the kmeans algorithm method [5].

Data mining clustering is one of the techniques in data mining used in this research to group data into similar groups based on similar characteristics. The main purpose of data mining clustering is to find similarities in data on the field of work of graduates and the position of graduates who are made into various clusters [6].

Data Mining Stages

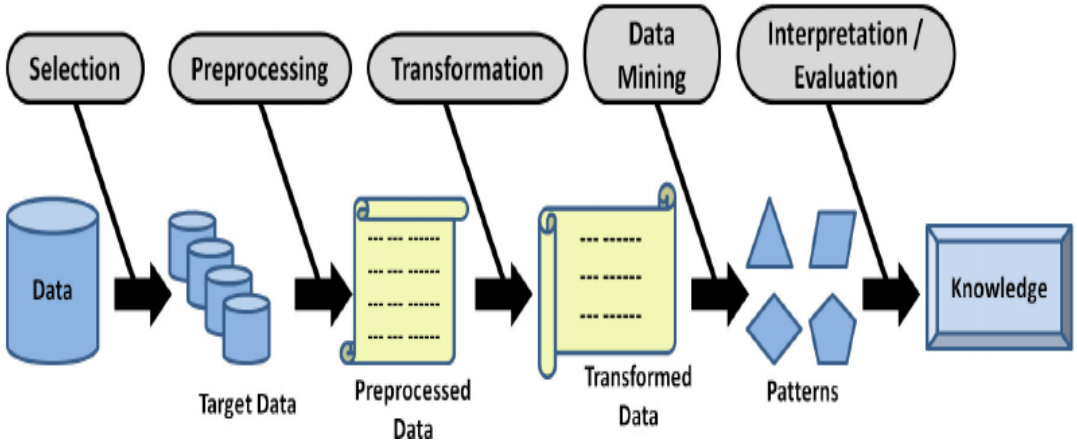
There are five stages that are passed in data mining, among others:

1. Data selection The first stage is the process of selecting data that will be used in the data clustering process. The dataset used is toddler nutritional status data at the Poto Tano District Health Center and determines the attributes that will be used for classification of 5,829 toddler data. The attributes to be used are name, age, gender, weight (BB), height (TB), and village. Data that has been selected or selected is then saved to be processed to the next stage.
2. Pre-processing / Cleaning Next, the cleaning process is the stage of cleaning irrelevant data to remove duplicate data. Cleaning is done on double data, empty data and eliminating unused data, and correcting data errors. Data cleaning is done manually using Microsoft Excel.
3. Data transformation (data transformation) Data is combined or transformed data that does not yet have a clear identity into a valid data form or relevant toddler data so that it is easily understood by RapidMiner tools for processing into data mining.
4. Data Mining The process of finding interesting patterns or information in selected data using classification and naïve bayes methods using RapidMiner tools.
5. Interpretation or evaluation The process of examining or reviewing data to a conclusion or the process of forming data output that is easy to understand and the results of determining stunting in toddlers produce relevant results and to produce an accuracy rate using classification [7].

Basic Concepts of Data Mining

A series of data mining processes, as for the concept, have the following stages:

1. Data Selection (selection)
   Performing data in the information mining stage in KDD begins.
2. Pre-processing (Data Selection)
   Enrichment of existing data with other data or information that is relevant and necessary for KDD, such as external data or information.
3. Transformation
   This is a creative process and depends on the type or pattern of information to be searched in the database.
4. 4.Data mining
   Analyzing interesting patterns from large amounts of data, then these data can be stored in the database.
5. Interpretation / Evalution
   Part of the KDD interpretation process includes checking whether or not the patterns or information found contradict pre-existing facts or hypotheses.The description in the Knowledge Discovery Database stage is as follows:



Method

The methods used to obtain secret messages and insert secret messages into images are the ROT13 and LSB methods.

Data Collection

The data collection conducted by the researcher is divided into several stages as follows:

1. Observation
   The researcher observes several images that will be used as trials on the application to be created.
2. Data Samples
   The researcher collects several samples that can be used as examples for this study, namely image data.
3. Literature
   The literature research conducted by the researcher aims to collect references used for this study.

K-Means Algorithm

K-Means Clustering is, K is intended as a constant number of desired clusters, Means in this case means the value of an average of a group of data which in this case is defined as a cluster, so K-Means Clustering is a data analysis method or data mining method that performs an unsupervised modeling process and is one of the methods that classify data with a partition system [10].

The steps to perform clustering with the K-Means method are as follows:

1. Determine the number of groups.
2. Allocate data into groups randomly.
3. Calculate the group center (centroid/average) of the data in each group. The centroid location of each group is taken from the average (mean) of all data values in each feature. If M denotes the number of data in a group, i denotes the i-th feature in a group, and p denotes the data dimension, then the equation to calculate the centroid of the i-th feature is equation 1. equation 1 is performed as many as p dimensions from i=1 to i=p.

Allocate each data to the nearest centroid/average. There are several ways that can be done to measure the distance of data to the center of the group, including Euclidean Distance measurement in the Euclidean distance space can be found using equation 2.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

The reallocation of data into each group in the K-Means method is based on the comparison of the distance between the data and the centroid of each existing group. The data is reallocated strictly to the group that has the centroid with the closest distance from the data.

Return to step 3, if there is still data that switches groups or if there is a change in the centroid value above the specified threshold value, or if the change in the value of the objective function used is still above the specified threshold value [11].

Results and Discussion

This section will show the results of the system design that has been built, namely the Data Mining application to classify daily sales transaction data.

Data Mining application to categorize daily sales transaction data. The results that will be displayed are the results of the interface display of the system that has been built and the results of system testing that has been carried out. The test method used is the K-Means Clustering method, which is one of the methods in Data Mining by clustering Indomaret's

daily sales transaction data. This system is expected to be able to provide useful, and relevant information in its information.

Interface Display Results

To test the correctness of the data mining results, Microsoft Visual Studio 2008 software is used.The following are the stages of the application of Data Mining on indomaret daily sales transaction data using the K-Means Clusterin method.

1. *Form* Login

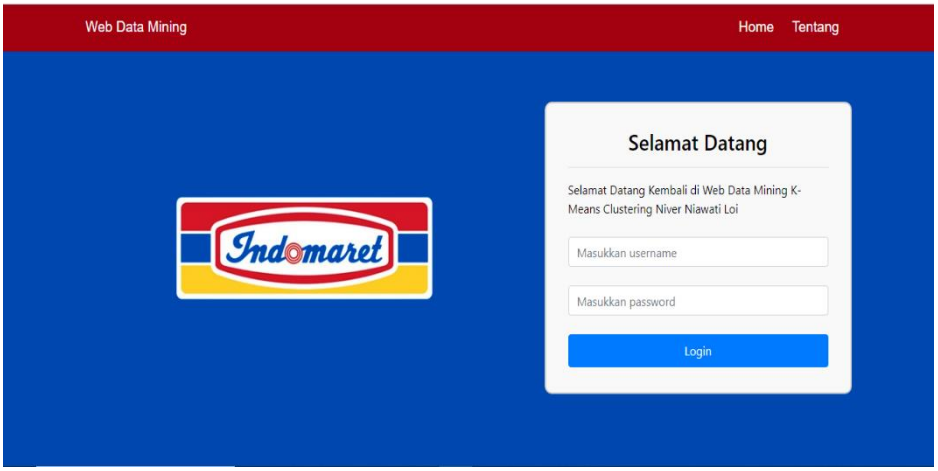


Figure 1: Login Form Display

2. Main display form

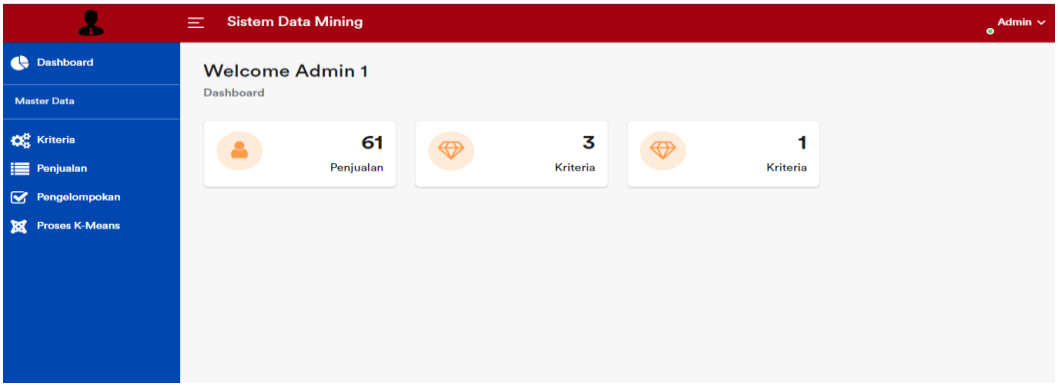


Figure 2: Tampilan Form Menu

3. Criteria Data Input Form

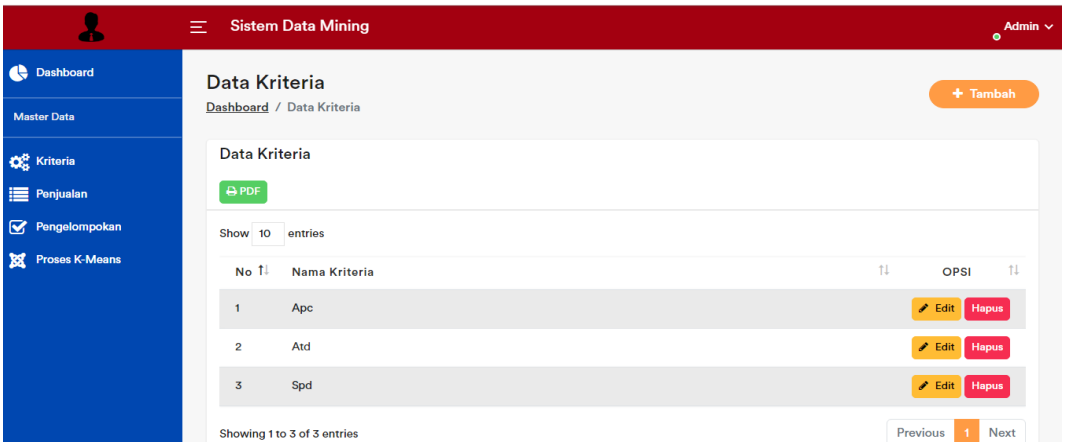


Figure 3: Criteria Data Input Form

4. **Sales Data Input Form**

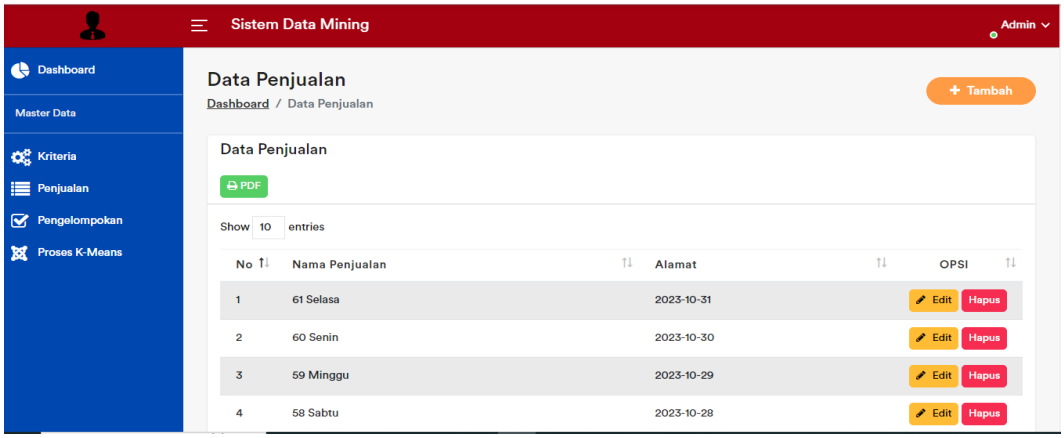


Figure 4: Sales Data Input Form

5. **Grouping Data Input Form**

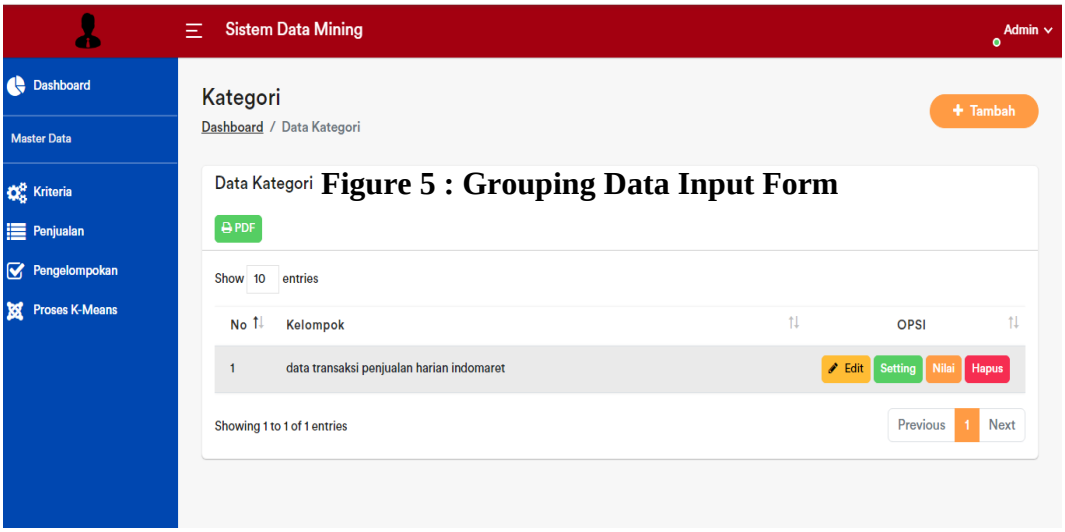


Figure 5 : Grouping Data Input Form

6. **K-Means Process Data Input Form**

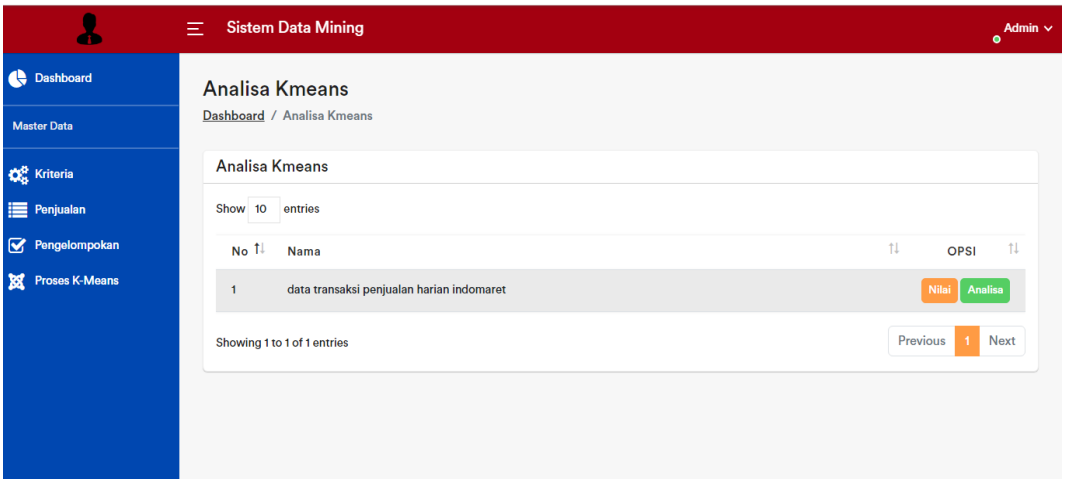


Figure 6: K-Means Process Data Input Form

System Testing

After implementing the system, the next step is to test the system that adopts the K-Means Clustering method. System testing is very important because it has several main objectives that help ensure the quality, reliability, and performance of the system being developed.

One of the tests on this developed system is using the Blackbox Testing method, a test conducted to observe the input and output results of the system. The following is testing by means of Blackbox Testing.

Table 1: *Black Box Testing*

No	Name Testing	Description	Validity
1	Click the Login Page	The application displays the image file selection	Valid
2	Click the Main Menu Page	The application performs Main Menu Page	Valid
3	Click DataCriteria page	The application performs Data Criteria	Valid
4	Click Sales Data Page	The application performs Sales Data	Valid
5	Click Grouping data page	The application displays Grouping Data Page	Valid
6	Click Data Page K-Means Process	The application performs Data Page K-Means Process	Valid

Trial Results

From the results of testing the system that has been designed, several advantages and disadvantages of the system can be identified and found. The following are the advantages and disadvantages of the system that has gone through the testing stage.

System Advantages

The advantages of this system are as follows:

1. This system can determine the grouping of transaction data which is very useful for knowing the daily sales profit at Indomaret.
2. In this system problem solving using the K-Means Clustering method, by juxtaposing the manual calculations contained in this study.

3. This system is made as minimalist as possible and makes it easy for users to operate.
4. This system is made web-based which can be accessed on a PC / laptop, tablet or smartphone device.
5. This system can run on any operating system

System Disadvantages

The disadvantages of this system are as follows:
1. Does not have a good security system, so it is still possible that this system can be hacked.
2. Users cannot interact directly with the admin, this is because the system does not provide messaging services..

Conclusion:

Based on the research and discussion of the previous chapters and the implementation of the program and testing based on data from the previous chapters, the following conclusions can be drawn:

1. How the implementation of data mining, especially using the K-Means Clustering algorithm, can be used to group sales transaction data at Indomaret, the implementation of data mining using the K-Means Clustering algorithm in the context of grouping daily sales transaction data at Indomaret can help Indomaret in knowing the sales profit every day.
2. What are the factors that influence the formation of grouping of Indomaret sales transaction data using K-Means Clustering, the factors that influence the formation of grouping of daily sales transaction data at Indomaret are the amount and type of data, such as the amount of daily sales transaction data in Indomaret.
3. How to design a system using the K-Means method at Indomaret, namely by collecting daily sales transaction data and determining the number of clusters to produce meaningful and actionable data groupings.

Bibliography

R. Mawarni, A. Silegar, P. Studi, and M. Informatika, “Pengolahan Data Mining Terhadap Penjualan,” vol. 04, no. 01, pp. 18–26, 2023.

H. Andika and S. W. Hati, “Analisis Perbandingan Kepuasan Pelanggan Antara Minimarket Indomaret Dengan Alfamart Di Kota Batam,” J. AKUNTANSI, Ekon. dan Manaj. BISNIS, vol. 6, no. 2, pp. 119–134, 2018, doi: 10.30871/jaemb.v6i2.651.

G. Hasan, D. Wistiasari, T. G. Hasvia, N. A. De Utami, and G. Aulia, “Analisis Penerapan Manajemen Operasional : Managing Quality pada Indomaret,” J. Minfo Polgan, vol. 12, no. 1, pp. 401–410, 2023, doi: 10.33395/jmp.v12i1.12414.

R. Dimilna, “Pengaruh Green Marketing Terhadap Minat Beli Konsumen Pada Produk Ecobag Di Indomaret (Studi Pada Indomaret Nusa Indah Kota Jambi),” Repository.Uinjambi.Ac.Id, vol. 1, no. 2, 2022.

N. Afiasari, N. Suarna, and N. Rahaningsi, “Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means,” J. SAINTEKOM, vol. 13, no. 1, pp. 100–110, 2023, doi: 10.33020/saintekom.v13i1.402.

Vrantika Br Samosir, Agung Mulyo Widodo, Nizirwan Anwar, Binastya Anggara Sekti, and Nixon Erzed, “Identifikasi Outlier Menggunakan Teknik Data Mining Clustering Untuk Analisis Data Tracer Study Pada Fakultas Ilmu Komputer Universitas Esa Unggul,” IKRA-ITH Inform. J. Komput. dan Inform., vol. 8, no. 1, pp. 162–174, 2024, doi: 10.37817/ikraith-informatika.v8i1.3211.

Y. Mulyanto, F. Idifitriani, A. Wati, U. T. Sumbawa, D. Mining, and K. P. Tano, "Vol 7 No 2 , September 2024 KLASIFIKASI DATA MINING UNTUK PENENTUAN STUNTING," vol. 7, no. 2, pp. 119–125, 2024.

Haris Kurniawan, Sarjon Defit, and Sumijan, "Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," J. Appl. Comput. Sci. Technol., vol. 1, no. 2, pp. 80–89, 2020, doi: 10.52158/jacost.v1i2.102.

A. O. P. Dewi, "Big Data di Perpustakaan dengan Memanfaatkan Data Mining," Anuva J. Kaji. Budaya, Perpustakaan, dan Inf., vol. 4, no. 2, pp. 223–230, 2020, doi: 10.14710/anuva.4.2.223-230.

Fina Nasari and S. Surya Darma, "Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru," Semin. Nas. Teknol. Inf. dan Multimed. 2015, pp. 73–78, 2015.

W. M. P. Dhuhita, "Clustering Menggunakan Metode K-Means Untuk," J. Inform., vol. 15, no. 2, pp. 160--174, 2016.

N. Sitohang, "Jurnal Sains Informatika Terapan (JSIT)," Penerapan Data Min. Untuk Peringatan Dini Banjir Menggunakan Metod. Klastering K-Means, vol. 2, no. 1, pp. 16–20, 2023.

H. D. Yunita, R. Okida, T. Winarko, and H. Sukri, "IMPLEMENTASI APLIKASI MONITORING INVENTARIS SPAREPART ELEKTRONIK BTS PADA PT . NEC INDONESIA," vol. 1, no. 2, pp. 57–64, 2024.

P. Studi, S. Informatika, and U. Raharja, "1 , 2 , 3," vol. 12, no. 2, pp. 47–56, 2023.

A. Josi, "Penerapan Metode Prototyping Dalam Membangun Website Desa (Studi Kasus Desa Sugihan Kecamatan Rambang)," Jti, vol. 9, no. 1, pp. 50–57, 2017.

M. Romzi and B. Kurniawan, "Implementasi Pemrograman Python Menggunakan Visual Studio Code," J. Inform. Dan Komput., vol. 11, no. 2, pp. 1–9, 2020, [Online]. Available: www.python.org

M. Suhartanto, "Pembuatan Website Sekolah Menengah Pertama Negeri 3 Delanggu Dengan Menggunakan Php Dan MySQL," J. Speed-Sentra Penelit. Enginerring dan Edukasi, vol. 4, no. 1, pp. 1–8, 2013.

M. Ahmia and H. Belbachir, "p, q-Analogue of a linear transformation preserving log-convexity," Indian J. Pure Appl. Math., vol. 49, no. 3, pp. 549–557, 2018, doi: 10.1007/s13226-018-0284-5.

C. Candra, A. Prabowo, and S. Suratno, "Analysis and Design of the Bpk Ri Library Information System Representative of Central Kalimantan Province Android Based," J. Sains Komput. dan Teknol. Inf., vol. 6, no. 1, pp. 72–77, 2023, doi: 10.33084/jsakti.v6i1.6056.

Sari, I.P., Al-Khowarizmi, A., Ramadhani, F., & Sulaiman, O.K. (2023). Implementation of the Selection Sort Algorithm to Sort Data in PHP Programming Language. Journal of Computer Science, Information Technology and Telecommunication Engineering 4 (1), 377-381

Ichsan, A., Al-Khowarizmi, A., & Azhari, M. (2024). Implementation of The Sales and Purchase Program Application Using the Rapid Application Development Model Web Based. Tsabit Journal of Computer Science 1 (1), 27-34

Sari, I.P., & Batubara, I.H. (2021). User Interface Information System for Using Account Services (Joint Account) WEB-Based. International Journal of Economic, Technology and Social Sciences (Injects) 2 (2), 462-469

Ramadhani, F., & Sari, I.P. (2021). Pemanfaatan Aplikasi Online dalam Digitalisasi Pasar Tradisional di Medan. Prosiding Seminar Nasional Kewirausahaan 2 (1), 806-811

Sari, I.P., & Alfarisi, F. (2024). Perancangan Sistem Aplikasi Pendataan Membership Gym Menggunakan Metode Unified Software Development Process (USDP) Berbasis Web. Hello World Jurnal Ilmu Komputer 3 (1), 37-48

Sari, I.P. (2020). Implementasi Pembayaran SPP Berbasis WEB Pada Sekolah Menengah Pertama (SMP) Muhammadiyah Kota Medan. Jurnal Pengabdian Barelang 2 (03), 11-14

Habib, T.A., Azly, R., Irza, M.A., & Prasetya, I. (2024). User Interface Design for the Orca Music Player Mobile Application. Tsabit Journal of Computer Science 1 (1), 18-26

Sari, I.P., Batubara, I.H., Ramadhani, F., & Wardani, S. (2022). Perancangan Sistem Antrian pada Wahana Hiburan dengan Metode First In First Out (FIFO). Sudo Jurnal Teknik Informatika 1 (3), 116-123

Ramadhani, F., Satria, A., & Sari, I.P. (2022). Aplikasi internet berbasis website sebagai E-Commerce penjualan komponen sport car. Blend Sains Jurnal Teknik 1 (2), 69-75

Sari, I.P., Ramadhani, F., Satria, A., Apdilah, D., & Basri, M. (2023). Rancangan UI/UX Aplikasi Analytics pada Toko Online Wao Sneakers Menggunakan Figma Berbasis Mobile. Factory Jurnal Industri, Manajemen dan Rekayasa Sistem Industri 1 (3), 93-101

Sari, I.P., Al-Khowarizmi, A., & Batubara, I.H. (2021). Implementasi Aplikasi Mobile Learning Sistem Manajemen Soal dan Ujian Berbasis Web Pada Platform Android. IHSAN: JURNAL PENGABDIAN MASYARAKAT 3 (2), 178-183

Sari, I.P., & Ramadhani, F. (2021). User Interface Prototype Using User Centered System Design Method in Motorvice Information System. 2021 International Conference on Computer Science and Engineering (IC2SE) 1, 1-6

Ramadhani, F., Sari, I.P., & Satria, A. (2024). Perancangan UI/UX Surat Keterangan Waris dalam Pengembalian Dana Haji Berbasis Web. Blend Sains Jurnal Teknik 2 (3), 198-203

Sari, I.P., Hariani, P.P., Satria, A., & Manurung, A.A. (2023). Rancang Bangun Sistem Informasi Pengelolaan Arsip Materi Ajar Berbasis Web untuk Guru MAS Darul Falah. Wahana Jurnal Pengabdian kepada Masyarakat 2 (2), 59-65

Sari, I.P., Syafii, R., Lubis, D.F.,, Setyadi, A., & Nasution, P. (2022). Pemanfaatan fasilitas google dalam perkuliahan di fakultas teknologi informasi. Blend Sains Jurnal Teknik 1 (2), 107-113

Ramadhani, F., & Sari, I.P. (2021). Improving the Performance of Naïve Bayes Algorithm by Reducing the Attributes of Dataset Using Gain Ratio and Adaboost. 2021 International Conference on Computer Science and Engineering (IC2SE) 1, 1-5

Sari, I.P., Sulaiman, O.K., Al-Khowarizmi, A., & Azhari, M. (2023). Perancangan Sistem Informasi Pelayanan Masyarakat pada Kelurahan Sipagimbar dengan Metode Prototype Berbasis Web. Blend Sains Jurnal Teknik 2 (2), 125-134

Sitompul, D.N., Rahmatika, A., & Sari, I.P. (2023). Application of The Sales and Purchase Program Using The Rapid Application Development Model. Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal 4 (1), 6-16

Sari, I.P., Ramadhani, F., Satria, A., & Apdilah, D. (2023). Implementasi Pengolahan Citra Digital dalam Pengenalan Wajah menggunakan Algoritma PCA dan Viola Jones. Hello World Jurnal Ilmu Komputer 2 (3), 146-157

Sari, I.P., Sulaiman, O.K., Ramadhani, F., & Satria, A. (2023). Perancangan Sistem Manajemen Surat Berbasis Web Pada Kantor Camat Tano Tombangan Angkola. INCODING: Journal of Informatics and Computer Science Engineering 3 (2), 61-76

Guntur, S., Ichsan, A., & Sari, I.P. (2024). Designing a Web-Based Mail Management System at the Beringin Helvetia Sub-district Office. Altafani: Jurnal Pengabdian Masyarakat 1 (1)

Sari, I.P., Al-Khowarizmi, A., Jannah, A., Meuraxa, A.M., & Tanjung, M.I. (2023). Web-Based Offline Game Suit Design: A Model Overview. Journal of Computer Science, Information Technology and Telecommunication Engineering 4 (2), 389-394

Sari, I.P., Al-Khowarizmi, A., Sulaiman, O.K., & Apdilah, D. (2024). System Design for Ordering and Digitizing Website-Based Bus Tickets. Journal of Computer Science, Information Technology and Telecommunication Engineering 5 (1), 543-549

Indah Purnama Sari. Algoritma dan Pemrograman. Medan: UMSU Press, 2023, pp. 290.

Janner Simarmata Arsan Kumala Jaya, Syarifah Fitrah Ramadhani, Niel Ananto, Abdul Karim, Betrisandi, Muhammad Ilham Alhari, Cucut Susanto, Suardinata, Indah Purnama Sari, Edson Yahuda Putra. Komputer dan Masyarakat. Medan: Yayasan Kita Menulis, 2024, pp.162.

Mahdianta Pandia, Indah Purnama Sari, Alexander Wirapraja Fergie Joanda Kaunang, Syarifah Fitrah Ramadhani Stenly Richard Pungus, Sudirman, Suardinata Jimmy Herawan Moedjahedy, Elly Warni, Debby Erce Sondakh. Pengantar Bahasa Pemrograman Python. Medan : Yayasan Kita Menulis, 2024, pp.180

Zelvi Gustiana Arif Dwinanto, Indah Purnama Sari, Janner Simarmata Mahdianta Pandia, Supriadi Syam, Semmy Wellem Taju Fitrah Eka Susilawati, Asmah Akhriana, Rolly Junius Lontaan Fergie Joanda Kaunang. Perkembangan Teknologi Informatika. Medan: Yayasan Kita Menulis, 2024, pp.158

Indah Purnama Sari. Buku Ajar Pemrograman Internet Dasar. Medan: UMSU Press, 2022, pp. 300.

Indah Purnama Sari. Buku Ajar Rekayasa Perangkat Lunak. Medan: UMSU Press, 2021, pp. 228.