Sales System Using Apriori Algorithm to Analyze Consumer Purchase Patterns

Elfina Novalia[1]
Study Program
Information Systems
Faculty of Engineering and Computer Science,
Universitas Buana Perjuangan Karawang
elfinanovalia@ubpkarawang.ac.id

Apriade Voutama[2]
Study Program
Information Systems
Faculty of Computer Science,
Universitas Singaperbangsa Karawang
Apriade.voutama@staff.unsika.ac.id

Syahri Susanto[3]
School of oil Technic
Akademi Minyask & Gas Balongan
syahri28@gmail.com

‹β›

Abstract—Penelitian ini bertujuan untuk membuat sebuah sistem penjualan untuk mendapatkan data pesanan tepat waktu, tidak terlambat sampai dalam hitungan hari, dan data menjadi terstruktur. Serta mengembangkan solusi untuk mengolah data transaksi penjualan yang akan semakin banyak menggunakan algoritma apriori untuk mengetahui pola pembelian konsumen sehingga dapat menjadi output untuk pengambilan keputusan atau pengetahuan. Penelitian ini menggunakan metode kualitatif untuk memperdalam pemahaman tentang fenomena yang ada saat ini sedalam mungkin. Hal ini menunjukkan pentingnya kedalaman dan detail dari data yang dipelajari. Pengembangan sistem menggunakan metode waterfall karena sangat sesuai dengan kebutuhan sistem yang akan dibangun. Dari hasil penelitian, perhitungan sampel data transaksi dengan total 12 data pada tanggal 7-8 Agustus 2021 menggunakan alat Tanagra menghasilkan aturan asosiasi bahwa jika Anda membeli pusaran, Anda akan membeli Caraco dengan nilai support sebesar 58% dan nilai confidence 100%, memiliki nilai lift ratio sebesar 1,3 menyatakan bahwa kedua produk tersebut memiliki keterikatan yang kokoh satu sama lain. Diikuti oleh jika Anda membeli Faraco, Anda akan membeli pusaran. Jika Anda percaya pada kristal, Anda akan membeli arco yang memenuhi kriteria parameter yang ditentukan dengan nilai dukungan minimum 20% dan kepercayaan minimum 50%.

Keywords: Penjualan, Data Mining, Apriori Algorithms.

Abstract—This study aims to create a sales system to get order data on time, not too late to result in days, and the data becomes structured. As well as develop solutions to process sales transaction data which will increasingly use a priori algorithms to find out consumer buying patterns so that they can be output for decision making or knowledge. This study uses a qualitative method to deepen understanding of the phenomena currently happening as profoundly as possible. This shows the importance of depth and detail of the data studied. The system development uses the waterfall method because it fits perfectly with the needs of the system to be built. From the results of the study, calculating a sample of transaction data with a total of 12 data on August 7-8, 2021, using the Tanagra tools resulted in a rule association that if you buy a vortex, you will buy a Caraco with a support value of 58% and a confidence value of 100%, having a lift ratio value of 1.3 stated that the two products have a solid attachment to each other. Followed by if you buy Faraco, you will purchase a vortex. If you believe in a crystal, you will buy an arco that meets the specified parameter criteria with a minimum support value of 20% and minimum confidence of 50%.

Keywords: Sales, Data Mining, Apriori Algorithms.

I. INTRODUCTION

Technological developments from time to time continue to develop very quickly. Likewise, what happened in the industrial era 4.0, where we are currently in an age that can be made easier to find and get what information we need through cyberspace as if the world is in our own hands. Technology brings significant changes to its users. Technology has both positive and negative impacts. In the industrial world, technology has an essential role in maintaining and caring for important data owned by a company agency.

Distributor UD Bangun Persada is a subsidiary of PT Triton Paint located in Malang City, East Java. PT Triton Paint is a company engaged in paint production. Various kinds of paints produced include wall paint, iron paint and wood. Distributor Cat UD Bangun Persada, located in the Karawang area, is a subsidiary of the West Java main office in the Cirebon area. Currently, it has six employees consisting of 1 sales head, one admin, three sales and one courier. As well as having material shops that are members of the UD Bangun Persada Paint Distributor, this number will continue to increase from time to time. Areas that become distribution centres for UD Distributors build Persada such as Karawang, Purwakarta, Bekasi and other areas.

UD Bangun Persada Paint Distributor only markets paint products produced by PT TRITON PAINT. In addition, it does not market products from other companies. In the case of sales transactions, they still record in writing, namely by way of sales recording what is ordered by the consumer on the order form. Then the structure is given to the admin or hung on the shelf. After that admin inputs consumer orders, several problems are found, including the sales data being not on time, and sales data is still often carried by sales and not stored in a structured manner. From sales data that is increasingly piling up, a solution can be made to be processed as well as possible using apriori algorithm data mining to find out consumer purchasing patterns, namely the itemset pattern

to determine the attachment of one item pattern to another item that can produce information that can be used for decision making. decisions and gain knowledge.

II. METHOD

A. System

A system is a procedural network of interconnected ones that collectively perform operations or achieve certain goals [1]. The system is a procedure or interrelated elements that have input, process, and output for the system to achieve its goals [2].

B. Information

Information is data that has been classified, processed or interpreted for use in decision making. Information processing systems convert data into information or process unnecessary data to be useful to the recipient[3]. Information is data that is processed in a format that is more useful and meaningful to the recipient, and data is a source of information that describes actual events [4]

C. Sales

Sales are receipts obtained from the delivery of merchandise or from the delivery of services on the stock exchange as consideration items, namely in the form of cash, cash equipment or other assets [5]. Sales is the gathering of a buyer and seller with the aim of exchanging goods and services based on valuable considerations, such as money considerations [6].

D. Apriori Algorithm

The Apriori algorithm is a method for finding the pattern of relationships between one or more elements in a data set, the Apriori algorithm is known as the market basket. The a priori algorithm can understand the buying patterns of consumers in the case that there is a 50% chance that consumers will buy goods A and B and then goods and C [7]. Apriori is a class algorithm that helps learn association rules. It works against transactions. The algorithm tries to find a common subset of a data set. A minimum threshold must be met for the association to be confirmed[8].

E. Tanagra

Tanagra is free software for academic and research purposes. This research involves several methods in data mining ranging from data exploration analysis, statistical learning, machine learning to databases [9]. Tanagra is one of the data mining software in which several data mining methods are provided, starting from exploring data analysis, statistical learning, machine learning and databases. Unlike most data mining software, tanagra is an open source based software where everyone can access the source code, and add their own algorithms, as long as he agrees and conforms to the software distribution license.[10].

F. Waterfall

The waterfall model is the simplest SDLC (Software Development Life Cycle) model. This model is only suitable for software development with specifications that do not change [1]. Waterfall provides a sequential or sequential software lifeflow approach starting from the analysis, design, coding, testing and support stages[11].

G. Data collection techniques

This method is compiled based on the results of the analysis of the research model that will be used, the results of the selection of system development, the waterfall model is used [12]. Literature study is done by looking at books, journals and previous scientific works to learn and find out information related to the author's research. Observation or observation is one of the primary data collection techniques by directly observing an activity carried out [13]. Interviews are primary data collection techniques by directly face to face with the interviewee[13].
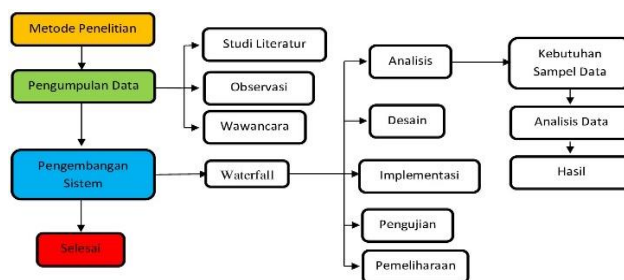
H. System development method

Figure 1 Research Flowchart

Analysis of the data to be searched, such as supporting data for system development and observing how the sales process and sales transaction data processing, as well as requiring sales transaction data samples for the a priori algorithm calculation process using Tanagra tools which can find out patterns of consumer buying associations, and references These calculations will be compared with the a priori algorithm calculations in the system. The need for transaction data samples with the aim of understanding the attributes on the sales form and selecting attributes for the purpose of data mining processes. In the data analysis stage, the sample of transaction data is calculated using the Tanagra tools with a minimum reference of 20% support and 50% minimum confidence. The result is the conclusion of the rule association that is formed from the calculation of transaction data using the Tanagra tools.

The design stage is the design of the system to be built. This design uses the UML (Unified Modeling Language) model, consisting of several steps, such as Use Case Diagrams, Activity Diagrams, Sequence, and Class Diagrams. Implementation from the design stage into coding in a programming language. The Programming language used is PHP Java and uses a MySQL database and the Codeigniter framework. Whitebox testing is focused on the internal system, namely the source code of the program [14]. Blackbox testing is done by testing system applications that involve users, which aims to find out the shortcomings of the application system that has been built. Maintenance At this stage, the care that has been developed is carried out. This treatment is to prevent errors found in the system carried out on the method according to the software requirements to keep it stable[15].

III. RESULTS AND DISCUSSION

A. Data Analysis

The data analysis process uses the Tanagra tools, in the early stages of determining a sample of transaction data. To find the rule association, determine the minimum support with 20% criteria while the minimum confidence is 50%.

Sampel Data		
id_transaksi	transaction_date	produk
TR07082101	2020-08-07	kristal,varaco
TR07082102	2020-08-07	varaco,arco
TR07082103	2020-08-07	varaco,kristal,vortex
TR07082104	2020-08-07	vortex,arco,varaco
TR07082105	2020-08-07	kristal,varaco,vortex
TR07082106	2020-08-08	arco,mexico,dextro
TR07082107	2020-08-08	vortex,arco,varaco
TR07082108	2020-08-08	varaco,vortex,mexico
TR07082109	2020-08-08	varaco,arco,vortex
TR07082110	2020-08-08	vortex,varaco,arco
TR07082111	2020-06-08	arco,dextro
TR07082112	2020-06-08	kristal,dextro

Figure 2 Sample Transaction Data

Before using the Tanagra tools, in the next stage, converting transaction data as shown above into binary format with the following results:

	A	B	C	D	E	F	G
1	No	kristal	vortex	varaco	dextro	mexico	arco
2	1	1	0	1	0	0	0
3	2	0	0	1	0	0	1
4	3	1	1	1	0	0	0
5	4	0	1	1	0	0	1
6	5	1	1	1	0	0	0
7	6	0	0	0	1	1	1
8	7	0	1	1	0	0	1
9	8	0	1	1	0	1	0
10	9	0	1	1	0	0	1
11	10	0	1	1	0	0	1
12	11	0	0	0	1	0	1
13	12	1	0	0	1	0	0

Figure 3 Tabular Format

After converting the data into binary format, the next step is to upload the binary format data into the Tanagra tool.
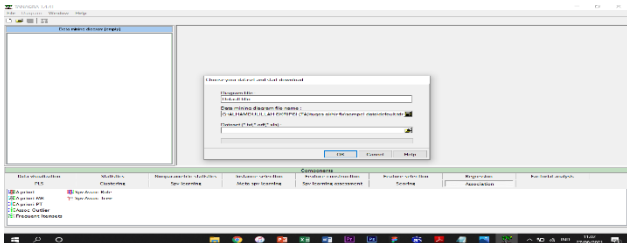
Figure 4 Tanagra Dashboard

Then in the next process click define status 1, select the attribute that will be processed for mining, then select ok.
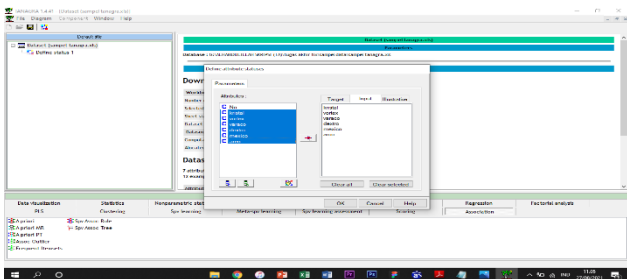

Figure 5 Define Status 1

Then at the bottom select the association menu, then drop frequent itemsets into define status 1. In frequent itemsets 1, right click, then select parameters. As has been determined in

the early stages of the parameters for a minimum support of 20%. Then select ok.
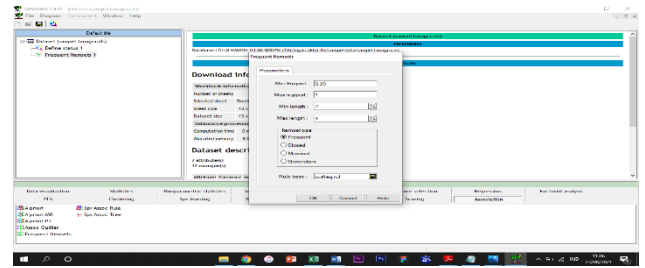

Figure 6 Frequent Itemsets 1

The next step is right click on Frequent Items 1. click execute, right click on frequent itemsets 1 then select view. then the itemset minimum support 20% will appear as follows.
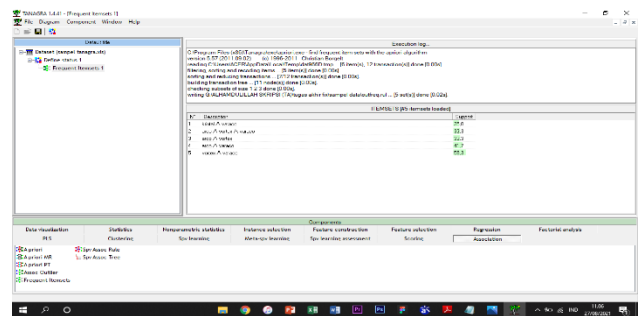

Figure 7 Results of 20% Support Itemsets

Then select the association menu again and select A priori, then drop into define status 1. right click on A priori, select parameters then input minimum support 20% and minimum confidence 50% then press ok.
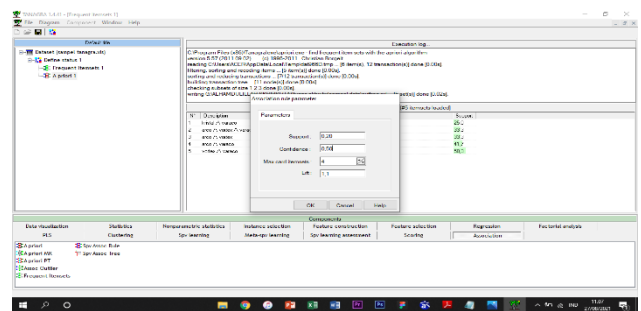

Figure 8 Input Support and Confidence Value

The next step right click on A priori. press execute, right click then view. Then the results of the rule association will appear with a minimum support of 20% and a minimum of 50% confidence as follows.

RULES

	Number of rules : 9				
N°	Antecedent	Consequent	Lift	Support (%)	Confidence (%)
1	"vortex=true"	"varaco=true" - "arco=true"	1,37143	33,333	57,143
2	"varaco=true" - "arco=true"	"vortex=true"	1,37143	33,333	80,000
3	"vortex=true" - "kristal=true"	"varaco=true"	1,33333	16,667	100,000
4	"vortex=true"	"varaco=true"	1,33333	58,333	100,000
5	"arco=true" - "vortex=true"	"varaco=true"	1,33333	33,333	100,000
6	"varaco=true"	"vortex=true"	1,33333	58,333	77,778
7	"varaco=true" - "kristal=true"	"vortex=true"	1,14286	16,667	66,667
8	"dextro=true"	"arco=true"	1,14286	16,667	66,667
9	"kristal=true"	"varaco=true"	1,00000	25,000	75,000

Computation time : 0 ms.

Figure 9 Results of the Rule Association

From the results of mining calculations using Tanagra tools with minimum support parameters of 20% and

minimum confidence of 50%, conclusions can be drawn that produce association rules. if you buy a vortex, you will buy a varaco with a support value of 58% and a confidence value of 100%. Having a lift ratio value of 1.3 indicates that the two products have a strong attachment to each other. Followed by varaco => vortex, crystal => arco that meets the predetermined parameter criteria.

B. System Design

The system design is the result of the implementation of the system analysis stage, in system modeling using UML (Unified Modeling Language) diagrams which consist of Use Case Diagrams, Activity Diagrams, Sequence Diagrams and Class Diagrams, the following are the results of the system design analysis.

a) Use Case Diagram

Use Case admin The admin use case describes the role of actors who have access to the system, such as having access rights to login, managing customer data, managing product data, managing transactions, managing transaction data for data mining processing purposes, managing mining processes to generate association rules, managing results. mining, reports and can manage user management. The sales use case describes the role of actors who have access to the system, such as having login access rights, and managing sales transactions for product orders from customers.

Use Case Sales coordinator describes the role of actor activities who have access to the system, and have login access rights. The sales coordinator can manage transaction data for mining processing purposes, manage the mining process to generate association rules, and manage reports.
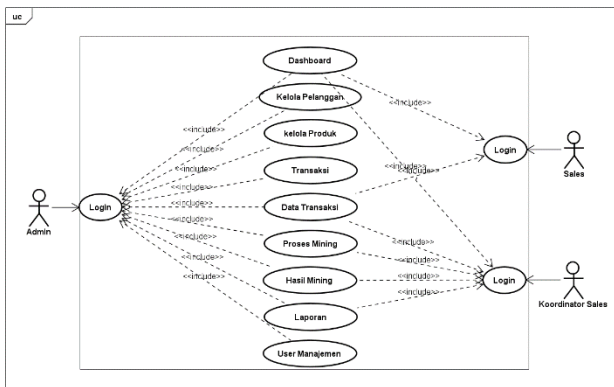

Figure 10 Use Case Diagram

b) Activity Diagram

This activity is carried out by the admin to process sales transaction data using the apriroi algorithm, the admin selects the a priori submenu and then the processing mining form appears. After that select the transaction date range to be processed then input the minimum support and minimum confidence then submit the process, then the system displays the association rule that has been processed by mining.

c) Sequence Diagram

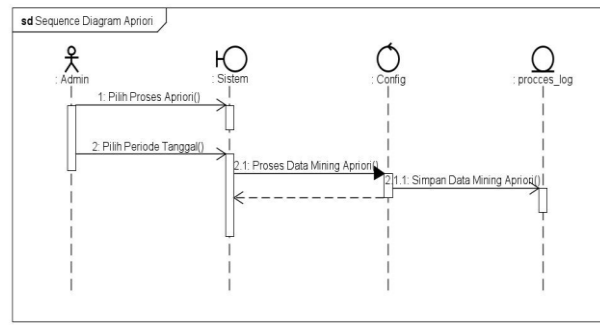This activity is for admins when doing the sales transaction data mining process.

Figure 11 Apriori Process Sequence Diagram

d) Class diagram

Class diagram Is a description of the system structure consisting of database attributes used in building a system, the following is a description of the class diagram.
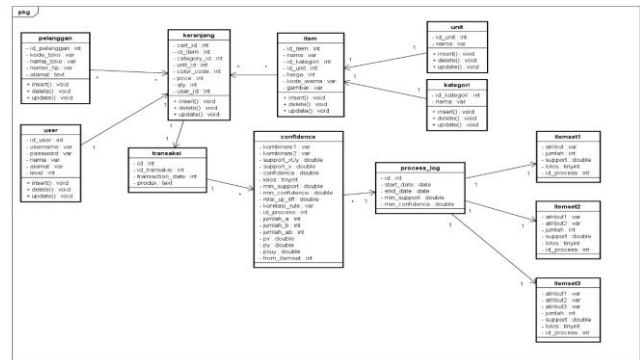

Figure 12 Class Diagram

A. System Implementation

The following is the association rule resulting from the system calculation.

Figure 13 Rule Association Counting System

The following is the association rule resulting from the calculation of the tanagra tools.
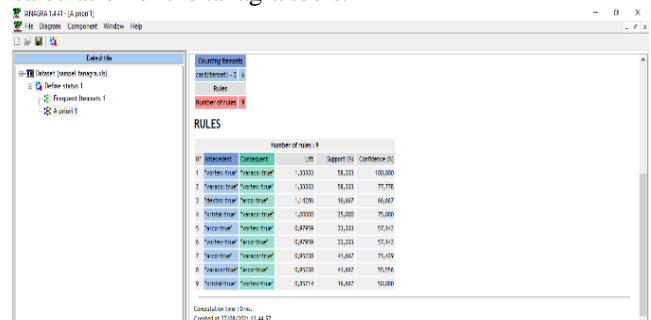

Figure 14 Rule Association Tanagra

Every user who has logged in according to their access rights will be immediately directed to the dashboard page.
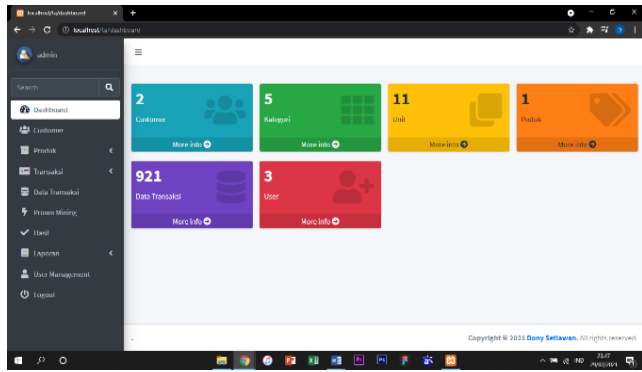

Figure 15 Dashboard Pages

The Transaction page is a process page for finding association rules, this page can be accessed by admin and sales coordinators.
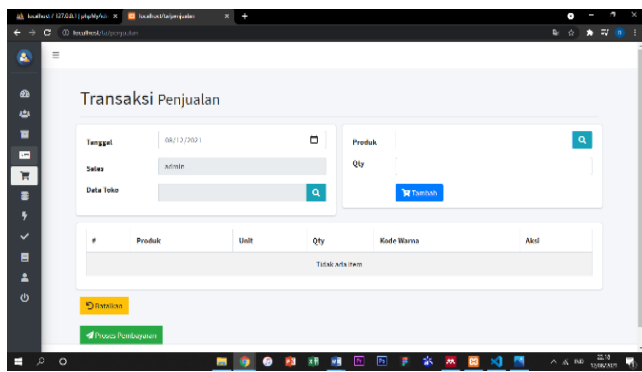

Figure 16 Sales Transaction Page

Mining Process page is a process page to find association rules, this page can be accessed by admin and sales coordinator.
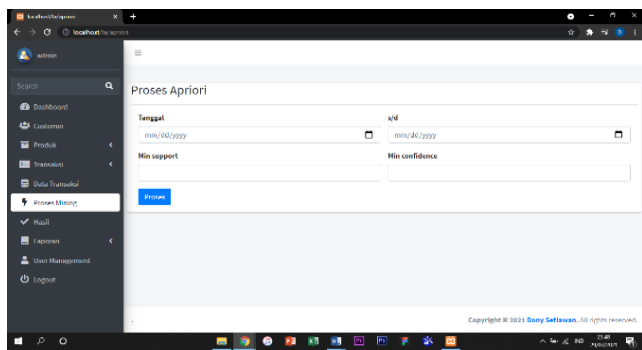

Figure 17 Mining Process Pages

The mining results page is a history of the results of the association rule search, this page can be accessed by admin and sales coordinators
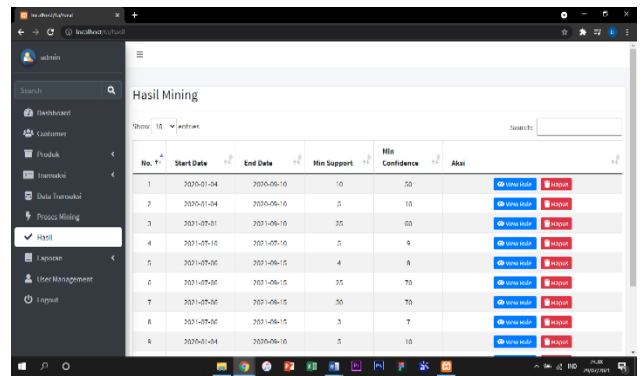

Figure 18 Mining Results page

IV. CONCLUSION

The results of calculations using the Tanagra tool with a total of 12 transaction data from August 7-8, 2021, produce a rule association: if you buy a vortex, you will buy a Caraco with a support value of 58% and a confidence value of 100%. A lift ratio value of 1.3 indicates that the two products have a solid attachment to each other. You are followed by Caraco => vortex, crystal => arco that meets the predetermined parameter criteria. After conducting an experimental calculation using the Tanagra tool and then comparing it with a system calculation and producing almost similar measures, the sales system using the a priori algorithm data mining follows the needs. So that UD Bangun Persada distributors can find out the product is purchasing patterns of their members in the sales application. Then the sales application generates sales data in real-time, and the sales data becomes structured so that you can view reports as needed.

REFERENCE

[1] A. Suryanto, "Rancang Bangun Sistem Informasi Pendaftaran Artis Berbasis Web Menggunakan Model Waterfall (Studi Kasus : Team Management Agensi)," *Jurnal Khatulistiwa Informatika*, vol. IV, no. 2, pp. 117–126, 2016.

[2] I. Laengge, H. F. Wowor, and M. D. Putro, "Sistem Pendukung Keputusan Dalam Menentukan Dosen Pembimbing Skripsi," *Jurnal Teknik Informatika*, vol. 9, no. 1, 2016.

[3] Fitri Ayu and Nia Permatasari, "perancangan sistem informasi pengolahan data PKL pada divisi humas PT pegadaian," *Jurnal Infra tech*, vol. 2, no. 2, pp. 12–26, 2018.

[4] R. Jermias, "Analisa Sistem Informasi Akuntansi Gaji Dan Upah Pada Pt. Bank Sinarmas Tbk. Manado," *Jurnal Riset Ekonomi, Manajemen, Bisnis dan Akuntansi*, vol. 4, no. 2, pp. 814–828, 2016.

[5] P. T. Smart, "Analisis Persediaan dan Penjualan Terhadap Arus Kas Operasi Pada," vol. 15, no. 1, 2021.

[6] E. Hikmawati, "PENJUALAN OBAT PADA APOTEK PUSAT DAN CABANG," no. 8, pp. 1–8.

[7] M. Syahril, K. Erwansyah, and M. Yetri, "Penerapan Data Mining Untuk Menentukan Pola Penjualan Peralatan Sekolah Pada Brand Wigglo Dengan

Menggunakan Algoritma Apriori," *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, vol. 3, no. 1, pp. 118–136, 2020.

[8]   I. Wahyudi, S. Bahri, and P. Handayani, "Aplikasi Pembelajaran Pengenalan Budaya Indonesia," vol. V, no. 1, pp. 135–138, 2019.

[9]   F. Rahmawati and N. Merlina, "Metode Data Mining Terhadap Data Penjualan Sparepart Mesin Fotocopy Menggunakan Algoritma Apriori," *PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic*, vol. 6, no. 1, pp. 9–20, 2018.

[10] H. Widayu, S. D. Nasution, N. Silalahi, and Mesran, "Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4.5," *Media Informatika Budidarma*, vol. Vol 1, No, no. 2, p. 37, 2017.

[11] S. Supriyati and D. M. Rizky, "Model Perancangan Sistem Informasi Akuntansi Budidaya Perikanan Berbasis SAK EMKM dan Android," *is The Best Accounting Information Systems and Information Technology Business Enterprise this is link for OJS us*, vol. 3, no. 2, pp. 301–315, 2018.

[12] B. Huda and B. Priyatna, "Penggunaan Aplikasi Content Management System (CMS) Untuk Pengembangan Bisnis Berbasis E-commerce," *Systematics*, vol. 1, no. 2, p. 81, 2019.

[13] A. L. Hananto and B. Priyatna, "Rancang Bangun Aplikasi Informasi Harga Produk," *TechnoXplore Jurnal Ilmu Komputer & Teknologi Informasi*, vol. 2, no. 1, pp. 10–20, 2017.

[14] S. Kasusdi, P. T. Rumah, and S. Padjadjaran, "Information System Journal," pp. 1–10, 2010.

[15] D. Rancadaka, "Aplikasi Penjualan Padi Berbasis Web Dengan Menggunakan Metode," vol. 1, no. 1, pp. 29–33, 2020.