Journal of Engineering, Electrical and Informatics

E-ISSN: 2809-8706

P-ISSN: 2810-0557

Implementation of Data Mining Algorithm C4.5 (Decision Tree) Product Distribution at PT.XYZ

Wulan Dari [1]*, Raden Aris Sugianto [2]

[1] Universitas Potensi Utama; ulandari2796@gmail.com
[2] Institut Teknologi Sawit Indonesia; ra.sugianto@itsi.ac.id

* Corresponding Author: Wulan Dari

Abstract . PT. XYZ in its distribution is still carried out by using the sales and the results of sales or distribution are recorded in a ledger about how many cigarette products are sold, as well as biodata from consumers. Based on the results of recording in the field will be recapitulated again at the office by using Microsoft Excel this is very difficult in the process of data integration, manufacturing, suppliers, retailers, sales and distribution of products. One of the things that can support the efficiency of production activities is to set the layout of the production machines owned. Material handling process, space utilization, distance between departments, production process flow can be more efficient if good layout planning is done. Supply Chain Management (Supply Chain Management) is a field of study that lies in the efficiency and effectiveness of the flow of cigarettes, information, and money flows that occur simultaneously so that it can unite the Supply Chain Management with the parties involved. The results of this study are a distribution information system that can be used to record all sales transactions. In order to make it easier for the data distribution section manager to recap and know the history of cigarette distribution better and complete reporting.

Keywords : Cigarettes, distribution, monitoring, PT. XYZ, Supply Chain Management

Received: April 16, 2025
Revised: April 30, 2025
Accepted: June 23, 2025
Published : June 24, 2025
Curr. Ver.: June 24, 2025

1. Background

In its daily operational development, it experiences many complexities. One of the departments that experiences this is the logistics department whose daily work is to combine raw material inventory data, raw material distribution, and production data. The data is collected at the end of working hours, making it difficult. This is done manually, so you can imagine how difficult it is if the data consists of thousands of data and requires a long process. This problem prompted PT. XYZ. to build Information Technology.

Data Mining is the process of finding patterns or finding interesting information from selected data using certain techniques or methods. The techniques used in data mining vary widely. The selection of the appropriate method or algorithm is highly dependent on the objectives and the overall KDD process [1].

Knowledge Discovery in Database (KDD) is as follows: Knowledge Discovery in Database (KDD) is a multi-level, non-trivial, interactive and iterative process for identifying understandable, valid, novel and potentially useful patterns from very large data sets [2].

Data mining is one part of the overall process for new unknown images, rules, and information in a fairly large amount of data, which includes data such as data cleaning, data transformation, data mining, image evaluation, and knowledge presentation, which is called the KDD process as follows:

1. Data Selection: Data is selected for processing and data is also mined for use when the KDD process begins.

DOI : https://doi.org/10.55606/jeei.v5i2.4540

2. Preprocessing: S before The process is continued in KDD , so it is necessary to carry out a cleaning process, namely a data cleaning process that is useful for removing unused data and avoiding duplicate data.
3. Transformation: A coding process on data that will be used , for data suitable for the data mining process. The coding process in KDD is creative work and depends on the type or image of information to be searched in the database.
4. Data mining: Searching for images or accurate information in selected data using certain techniques or methods.
5. Interpretation/Evaluation: The image results from the data mining process need to be displayed in a form that is easy for the user to understand .

The C4.5 algorithm is an algorithm used to form a decision tree. Decision trees are a very powerful and well-known classification and prediction method. The decision tree method transforms very large facts into decision trees that represent rules. [3].

In this study, a product distribution pattern search information system will be built at PT. XYZ so that it can help sellers to find out what cake equipment is often purchased together.

PT. XYZ. chooses to use ERP (enterprise resource planning) from Oracle, SDLC (system development life cycle), phasing strategy. With the Information Technology built by PT. Gudang Garam Tbk. this company can solve existing problems and is able to find out if there are problems that arise in every sector or part of this company quickly and can find answers to these problems and be able to fix them immediately.

The problem faced by PT. XYZ is the difficulty of predicting marketing patterns and how to face increasingly tight business competition, is an obstacle faced by PT. XYZ so that management must be able to determine the distribution pattern of cigarette products, so that management can make decisions when they find out the marketing image of the product that is not in demand by customers or consumers.

2. Research Methods

2.1 Data Mining

Data mining is the process of data mining or the process of finding new data by looking for certain patterns and rules from large data. Data mining is also defined as knowledge discovery in database (KDD), which is an activity that includes collecting, using data, history to find regularities, patterns or relationships in large data sets. Data mining is defined as the process of finding patterns in data. This process is usually automatic or semi-automatic. The patterns found must have meaning and benefits, usually economic benefits and the data needed in large quantities. (Heroe Santoso, et al., 2016)

Data mining is the process of extracting information from a collection of data. large through the use of algorithms and techniques involving the fields of statistics, machine learning, and database management systems. Association analysis or association rule mining is a data mining technique for determining association rules between a combination of items (Khairul, 2015).

2.2 C4.5 Algorithm (Decision Tree)

The C4.5 algorithm is an algorithm used to form a decision tree. Decision trees are a very powerful and well-known classification and prediction method. The decision tree method transforms very large facts into decision trees that represent rules. Rules can be easily understood. And they can also be expressed in the form of database languages such as Structured Query Language to search for records in a particular category [3]. The C4.5 algorithm method for building a decision tree is:

a. Select the attribute to use as the root

b. Create a branch for each value.

c. Divide the cases into a branch

d. Repeat the process for each branch until all cases on the branch have the same class.

There are several stages in creating a decision tree with the C4.5 algorithm [5].

1. Prepare training data. This data is taken from previously existing data and has been grouped into certain classes.

2. After that determine the root of the tree. Choose the root of the attribute, the way is to calculate the gain value of all attributes, the first root is the highest gain value. Before determining the gain value, first calculate the entropy value.

3. Research Methods

1. Existing System Analysis

In completing this research , the author used 2 (two) study methods, namely :

1. Field Study

Field Study is a research conducted by collecting data and information obtained directly from the group and directly observing tasks related to cigarette sales marketing at PT. XYZ (*Observation*).

2. *Library* Research

As a reference for making a thesis, it will be taken from a literature study. Literature studies can be done from references to books on *data mining* , journals or *textbooks* that can be obtained in libraries or from sources on the internet.

In developing the system, the author uses the *waterfall paradigm* . The *waterfall method* has the following stages:

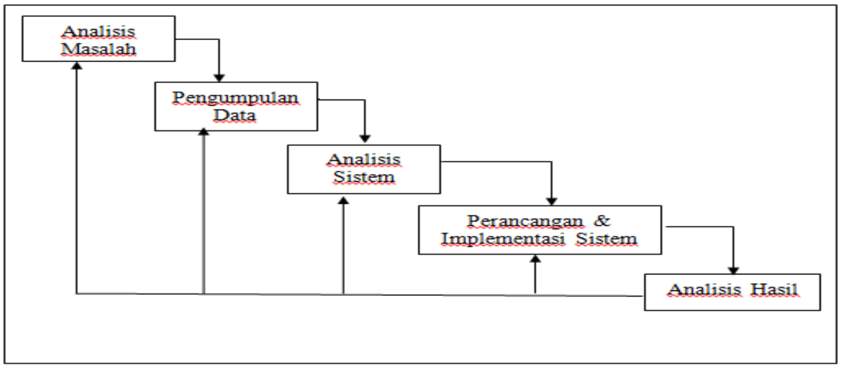


Figure 1 . *Waterfall*

1. Needs Analysis

The stages will be carried out by studying the problems found in the Sinar Raya store, determining the scope of the problem, studying several journal literature, textbooks and books and analyzing the required data.

a. Field study

This is a technique that is carried out by collecting data by conducting direct research on the research object and collecting data through:

b. Observation

Is collecting data by conducting direct inspections, reviewing and analyzing existing procedures.

2. Data Collection

a. Field Research

b. Library Research

3. System Analysis

At the system analysis stage, the needs required in building the application will be determined, so that it runs according to plan. The main purpose of the system analysis stage is to find out the capability requirements or criteria that must be met by the system so that the wishes of the system users can be realized.

4. Results And Discussion

3.1. Application of the C4.5 (Decision Tree) Method

The data used to analyze cigarette sales patterns that are definitely sold is transaction data.

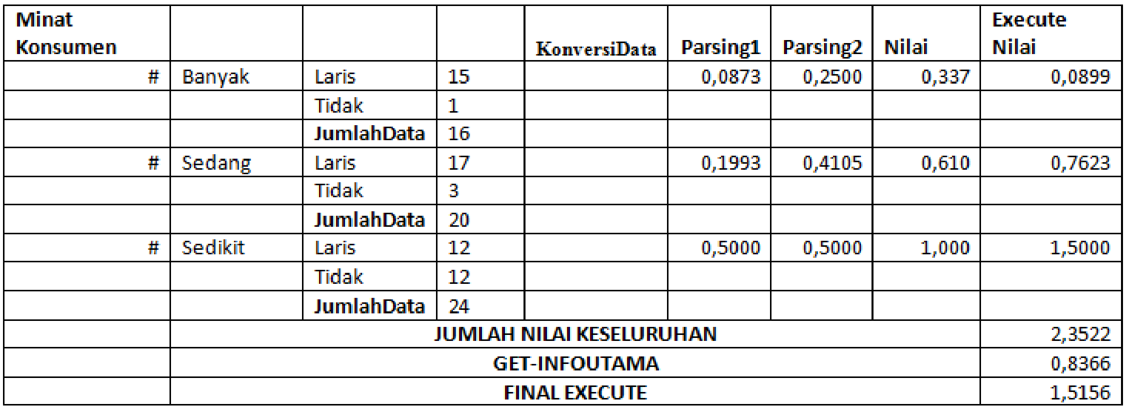
Cigarette Sales Data

No	Most Sold	Consumer Interest	Price	Prediction
1	Lots	Lots	Cheap	Popular
2	Lots	Lots	Cheap	Popular
3	Lots	A little	Cheap	No
4	A little	A little	Expensive	No
5	A little	Currently	Cheap	Popular
6	Currently	Currently	Cheap	No
7	Currently	Currently	Expensive	Popular
8	Currently	A little	Expensive	No
9	A little	A little	Cheap	No
10	A little	Currently	Cheap	Popular
11	Lots	Lots	Expensive	Popular
12	Currently	Lots	Expensive	No
13	Currently	Currently	Expensive	Popular
14	A little	A little	Cheap	No
15	A little	Lots	Cheap	Popular
16	Currently	Currently	Cheap	Popular
17	Lots	Lots	Cheap	Popular
18	Lots	A little	Cheap	Popular
19	A little	A little	Cheap	No
20	A little	Lots	Expensive	Popular
21	Currently	Lots	Expensive	Popular
22	Currently	Lots	Cheap	Popular
23	Lots	A little	Cheap	Popular
24	A little	A little	Cheap	No
25	A little	Lots	Cheap	Popular
26	Lots	Lots	Cheap	Popular
27	Currently	Lots	Expensive	Popular
28	Lots	A little	Cheap	Popular
29	A little	A little	Cheap	No
30	A little	Lots	Cheap	Popular
31	Currently	Lots	Cheap	Popular
32	Currently	Lots	Expensive	Popular
33	Lots	A little	Expensive	Popular

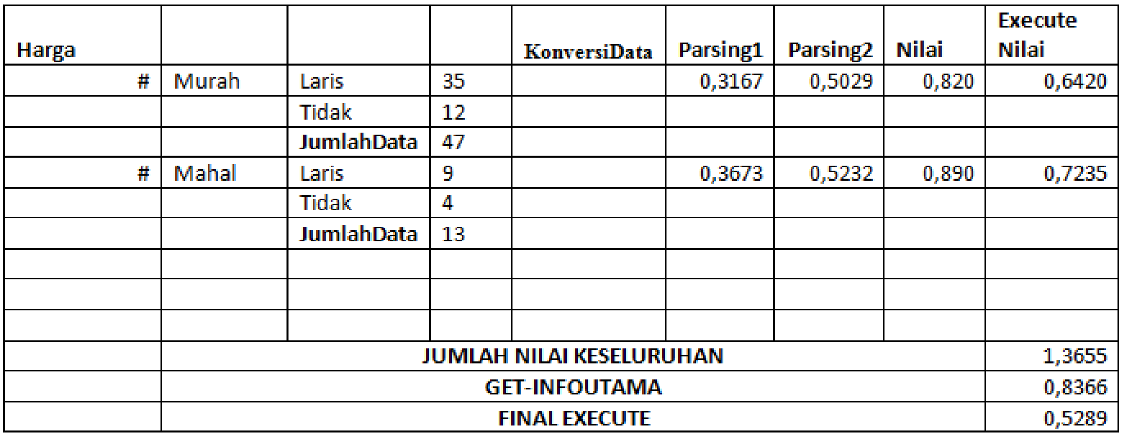
34	A little	A little	Expensive	No
35	A little	Lots	Cheap	Popular
36	Currently	Lots	Cheap	Popular
37	Currently	Lots	Cheap	Popular
38	Lots	A little	Cheap	Popular
39	A little	A little	Cheap	No
40	A little	Lots	Cheap	Popular
41	Currently	Lots	Cheap	Popular
42	Currently	Lots	Expensive	Popular
43	Lots	Currently	Cheap	Popular
44	A little	A little	Cheap	No
45	A little	Lots	Cheap	Popular
46	Lots	Lots	Cheap	Popular
47	Currently	Lots	Cheap	Popular
48	Currently	A little	Cheap	Popular
49	A little	A little	Cheap	No
50	A little	Lots	Cheap	Popular
51	Lots	Lots	Cheap	Popular
52	Lots	Currently	Cheap	Popular
53	Currently	A little	Cheap	Popular
54	A little	Currently	Cheap	No
55	A little	Lots	Cheap	Popular
56	Currently	Lots	Cheap	Popular
57	Currently	Lots	Cheap	Popular
58	Lots	A little	Cheap	Popular
59	A little	A little	Cheap	No
60	A little	Lots	Cheap	Popular

In the table above there are 60 assessments from 4 assessments that determine the cigarette sales pattern at PT.XYZ.

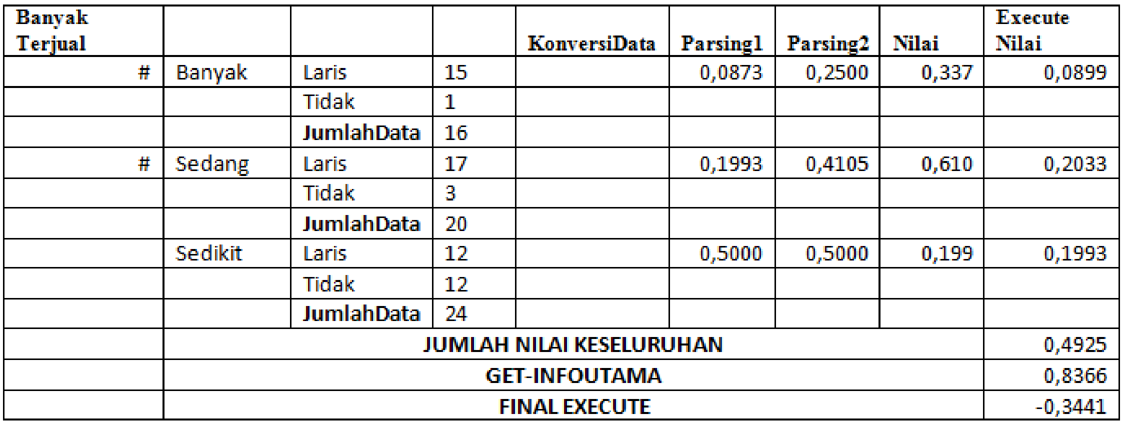
Table 2. Gain and Enthrophy Values



Minat Konsumen				KonversiData	Parsing1	Parsing2	Nilai	Execute Nilai
#	Banyak	Laris	15		0,0873	0,2500	0,337	0,0899
		Tidak	1					
		JumlahData	16					
#	Sedang	Laris	17		0,1993	0,4105	0,610	0,7623
		Tidak	3					
		JumlahData	20					
#	Sedikit	Laris	12		0,5000	0,5000	1,000	1,5000
		Tidak	12					
		JumlahData	24					
	JUMLAH NILAI KESELURUHAN							2,3522
	GET-INFOUTAMA							0,8366
	FINAL EXECUTE							1,5156



Harga				KonversiData	Parsing1	Parsing2	Nilai	Execute Nilai
#	Murah	Laris	35		0,3167	0,5029	0,820	0,6420
		Tidak	12					
		JumlahData	47					
#	Mahal	Laris	9		0,3673	0,5232	0,890	0,7235
		Tidak	4					
		JumlahData	13					
	JUMLAH NILAI KESELURUHAN							1,3655
	GET-INFOUTAMA							0,8366
	FINAL EXECUTE							0,5289



Banyak Terjual				KonversiData	Parsing1	Parsing2	Nilai	Execute Nilai
#	Banyak	Laris	15		0,0873	0,2500	0,337	0,0899
		Tidak	1					
		JumlahData	16					
#	Sedang	Laris	17		0,1993	0,4105	0,610	0,2033
		Tidak	3					
		JumlahData	20					
	Sedikit	Laris	12		0,5000	0,5000	0,199	0,1993
		Tidak	12					
		JumlahData	24					
	JUMLAH NILAI KESELURUHAN							0,4925
	GET-INFOUTAMA							0,8366
	FINAL EXECUTE							-0,3441

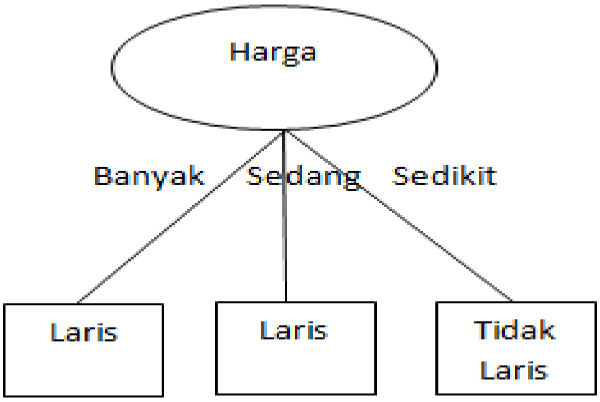


Figure 2. Manual Decision Tree

Next, testing with the specified tools, namely Rapidminer 5. The decision tree process in Rapidminer 5 starts from inputting

Conclusion

Based on the discussion of the previous chapters, the following conclusions can be drawn:

1. A computer application has been created that can store sales data and generate knowledge about buyers' interests, so PT .

2. In applying the C4.5 algorithm method, the C4.5 algorithm can be applied to produce knowledge regarding buyer interest in choosing a product.
3. By using *Visual Basic 2010 programming and using the SQL Server 2008* database , a system can be produced that can generate knowledge about buyer interest in choosing products.

Suggestion

For further development of the application of the C4.5 *data mining* algorithm in determining buyer interest in choosing products at PT.XYZ , the following suggestions can be given:

1. It is better if the system that has been created can be developed using other C4.5 methods.
2. It is better if the system that has been created can use the percentage of buyer interest in choosing products at PT. XYZ .
3. It is better if the system that has been created can be implemented using an *online-based system.*

Thank-you note

The author would like to express his gratitude to Universitas Potensi Utama for providing the means to gain knowledge so that the author can gain experience and teaching to be able to complete his education well.

Reference

[1.] *Analysis of Human Disease Data Patterns Caused by Cigarettes*, *COMICS (Conference on National Information and Computer Technology)*, vol. 1, no. 1, Bandung: Engineering Science, 2013.

[2.] H. Kurniawan, F. Agustin, Yusfriza, and K. Ummi, *Implementation of Data Mining in Prediction of Sales Chips With Rough Set Method*, 2018 6th International Conference On Cyber And IT Service Management (CITSM), Parapat, Indonesia, 2018.

[3.] Kusrini and E. T. Luthfi, *Data Mining Algorithms*. Yogyakarta: ANDI, 2009.

[4.] Nasari, F. and Darma, S., "Application of k-means clustering on new student admission data (case study: main potential university)," *SEMNASTEKNOMEDIA ONLINE*, vol. 3, no. 1, pp. 2-1, 2013.

[5.] Norsyaheera, A. W., Lailatul, F. A. H., Shahid, S. A. M., and Maon, S. N., "The Relationship Between Marketing Mix and Customer Loyalty in Hijab Industry: The Mediating Effect of Customer Satisfaction," *Procedia Economics and Finance*, vol. 37, pp. 366–371, 2016. https://doi.org/10.1016/S2212-5671(16)30138-1.

[6.] Rahmadya, T. H. and Herlawati, Prabowo, P. W., *Implementation of Data Mining with Matlab.*

[7.] Santoso, H., Hariyadi, I. P., and Prayitno, P., "Data Mining Analysis of Product Purchase Patterns Using the Apriori Algorithm Method," *Semnasteknomedia Online*, vol. 4, no. 1, pp. 3–7, 2016.

[8.] Sari, E. N., "Apriori Algorithm Analysis to Determine the Most Popular Clothing Brands in Medan Fashion Group Mode," *Pelita Informatika: Information and Informatics*, vol. 4, no. 3, 2013.

[9.] Ummi, K., "Data Mining Analysis in Car Spare Part Sales Using Apriori Algorithm Method (Case Study: At PT. Idk 1 Medan)," *CSRID (Computer Science Research and Its Development Journal)*, vol. 8, no. 3, pp. 155–164, 2016.

[10.] Urva, G. and Siregar, H. F., "UML Modeling of Cooking Oil E-Marketing," *ROYAL JOURNAL*, Edition 2, 2015.

[11.] Waruwu, F. T., Buulolo, E., and Ndruru, E., *Implementation of the Apriori Algorithm*, 2017.

[12.] Wulan Dari, "Implementation of Data Mining with Naive Bayes to Predict BOS Fund Recipients at School X," 2023.

[13.] Zulita, L. N., "Implementation of Selection Sort Method to Determine Achievement Value of Grade 3 and Grade 4 Students of SD Negeri 107 SELUMA," *Jurnal Media Infotama*, vol. 11, no. 1, 2015.