International Journal of Information System & Technology
Akreditasi No. 36/E/KPT/2019 | Vol. 5, No. 3, (2021), pp. 257-263

Application of the C4.5 Algorithm Determination of Depreciation Classification Volume of Fertilizer on Storage Warehouse

Yendrizal

Computer Engineering, AMIK KOSGORO

yendrizal70@gmail.com

Abstract

The problems faced by the fertilizer warehouse PT. Iskandar Muda currently shrinking the volume of fertilizer for storage in warehouses, which is often reduced, is not yet known by the management. In reducing the volume of this fertilizer, the author wants to apply the C4.5 Algorithm to help in making decisions. From all the existing data will be classified and predicted, from the classification results will present the rules. So that it can be easily understood in natural language. Based on the application of the C4.5 Algorithm in the classification technique of fertilizer volume shrinkage on storage in the warehouse at the Pupuk Iskandar Muda warehouse, the results of this study show that 1) the C4.5 algorithm with the decision tree method can provide executive information and the system is used to describe processes related to depreciation classification. 2) The C4.5 algorithm is considered as an algorithm that is very helpful in classifying data because the characteristics of the classified data can be obtained clearly, both in the form of a decision tree structure and rules or rule-if-then. 3) Form a decision tree in providing recommendations for decision making.

Keywords: *C4.5 Algorithm, Decision tree, Data Mining*

1. Introduction

Data mining is an important step in the process of finding knowledge in databases with intelligent thinking patterns. Data mining deals with training knowledge data and unlabeled data [1]. To apply the concept of data mining, there are many ways that can be used, including using the concept of several algorithms that exist in data mining. The C4.5 algorithm is one of the algorithms used to classify data by forming a decision tree. The problem faced by the Iskandar Muda fertilizer warehouse at this time is the shrinking of the volume of fertilizer for storage in the warehouse which is often reduced is not yet known by the management. In reducing the volume of this fertilizer, the author wants to apply the C4.5 Algorithm to help in making decisions. From all the existing data will be classified and predicted, from the classification results will present the rules. So that it can be easily understood in natural language. Data mining is a term used to describe the discovery of knowledge in databases. Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from large databases [2]. According to the Gartner Group, data mining is a process of finding meaningful relationships, patterns, and tendencies by examining large sets of data stored in storage using pattern recognition techniques such as statistical and mathematical techniques [3]. In addition to the above definitions, several definitions are also given as listed below. In his journal Kusrini explains that data mining is the process of looking for interesting patterns or information in selected data using certain techniques or methods [4]. From the definitions that have been presented, the important things related to data mining are:

a) Data mining is an automated process of existing data.
b) The data to be processed is in the form of very large data.

ISSN : 2580-7250

c) The purpose of data mining is to find relationships or patterns that may provide useful indications.

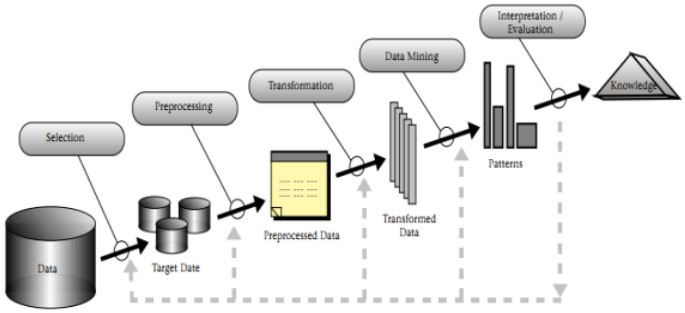

Figure 1. Stages of Knowledge Discovery in Database

Data mining can be grouped based on the tasks that can be done, namely [5]:

a) Description
Sometimes researchers and analysts simply want to try to find a way to describe the patterns and trends contained in the data. Thus descriptions of patterns and trends often provide possible explanations for a pattern or trend.

b) Estimate
Estimation is almost the same as classification, except that the estimation target variable is more numerical than categorical. The model is built using a complete record that provides the value of the target variable as the predicted value. Furthermore, in the next review, the estimated value of the target variable is made based on the value of the predictive variable.

c) Prediction
Prediction is almost the same as classification and estimation, except that in predicting the value of the results will be in the future. Some of the methods and techniques used in classification and estimation can also be used (for appropriate circumstances) for prediction.

d) Classification
In the classification, there are target categorical variables. For example, income classification can be separated into three categories, namely high income, medium income, and low income.

e) Clustering
Clustering is grouping records, observations, or observing and forming classes of objects that have similarities. Cluster is a collection of records that have similarities with one another and have dissimilarities with records in other clusters.

f) Association
The task of association in data mining is to find attributes that appear at one time. In the business world, it is more commonly called shopping cart analysis, which will be processed in the form of very large data.

Classification is a process to find a model or function that describes and separates a class from a data with the aim of being able to use the model or function to predict a class from a new object that has not been recognized [6]. Classification is a difficult task. happens very often in everyday life. Its basis involves dividing objects so that each is assigned to one of a number of mutually exclusive and mutually exclusive categories known as classes. The term 'mutually complete and exclusive' simply means that each object must be assigned to exactly one class, i.e. never more than one and never any class

at all. Classification consists of the process of examining one or more features of the presented new object and assigning it to one of the existing classes [7].

Classification results can be presented in various forms, namely classification rules (IF-THEN), decision trees or decision trees, mathematical formulas and neural networks. Many algorithms can be used in the formation of decision trees, including ID3, CART, and C4.5 . The C4.5 algorithm is one of the decision tree induction algorithms, namely ID3 (Interative Dichotomiser 3). ID3 was developed by J. Ross Quinlan. In the ID3 algorithm procedure, the input is a training sample, training table and attributes. The C4.5 algorithm is a development of ID3 [6]. PT. Iskandar Muda is one type of fertilizer company that is allowed to operate in Indonesia based on a fertilizer subsidy system to farmers. In the subsidized fertilizer system, PT. Iskandar Muda is a form of company whose management must be based on fertilizer subsidies for farmers and farmer groups. Legal rules regarding PT. Iskandar Muda refers to Government Regulation No. 12 of 2000. In the fertilizer distribution system to farmers and farmer groups. PT. Iskandar Muda is a company established to serve fertilizer needs for farmers, especially farmer groups that already have management and are registered as a cooperative with a legal entity (already recognized by the government). Where is the scope of work of PT. Iskadar Muda only serves designated areas such as Bukittinggi City, Agam Regency, Fifty Cities Regency, West Pasaman Regency and East Pasaman Regency in West Sumatra.

Depreciation is the reduced volume of fertilizer purchased by farmer groups to PT. Iskandar Muda after being accepted by the farmer groups. This is often complained of and submitted to the management of PT. Iskandar Muda. The shrinkage in the volume of this fertilizer stored in the warehouse is caused by the gancu, the weather, the place (Estapel) of storage, the length of time it is stored and so on. With this problem, it can be overcome by the management in making decisions.

2. Research Methodology

In this chapter, the research methodology will be discussed to identify problems, analyze problems and finally find solutions to problems in analyzing the shrinkage of fertilizer volume in storage warehouses at PT. Iskandar Muda. In the research methodology there is a sequence of frameworks that must be followed, the sequence of these frameworks is a description of the steps that must be passed so that this research can run well. The framework to be followed can be seen in Figure 2.

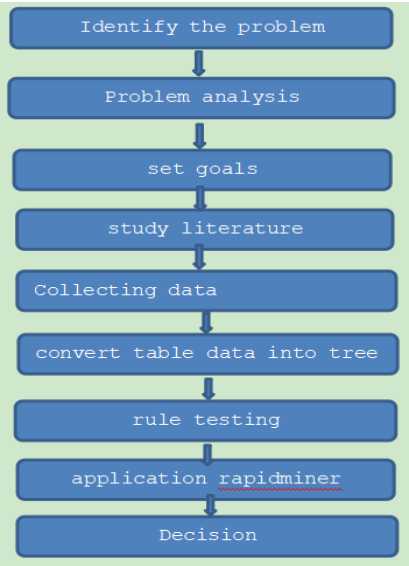

Figure 2. Framework

3. Results and Discussion

In this study the data collected are:

a) Depreciation analysis data.
b) Data recapitulation from research results obtained during field research.

Some of the attributes needed in determining the policy for shrinking the volume of fertilizer in storage warehouses are:

Table 1. Attributes of Fertilizer Volume Depreciation

No	Data Name	Attribute	Category
1	Storage	String	Moist, Hot, Dry
2	Storage Time	String	Day, Sunday, Month
3	Storage Arrangement	String	Normal, Procedure
4	Use Gancu	String	There's a mess, there's nothing

The above attributes can be explained as follows:

a) Storage

The storage area is a place that provides information about the shrinkage of the volume of fertilizer that occurs in the storage warehouse. Included in the storage area are:

1) Moist Location: this is a situation where fertilizer is stored in the storage room.
2) Location of heat: in the storage of direct or indirect sunlight to the fertilizer storage area.
3) Dry Location: a place to store fertilizers that are free from moisture and direct or indirect sunlight.

b) Storage Time

The retention time attribute is categorized as follows:

1) Day: the length of storage of fertilizer in the warehouse ranges from 1 day to 6 days.
2) Sunday: the storage time of fertilizer in the warehouse is only 1 week to 3 weeks.
3) Months: the storage time of fertilizer in the warehouse is only 1 month to 11 months.

c) Fertilizer Composition

The fertilizer composition attributes are categorized as follows:

1) Ordinary Arrangement: The arrangement of fertilizers in the storage warehouse is made parallel.
2) Arrangement of Procedures: the preparation of fertilizers in the storage warehouse is made regularly with a row in the same direction.

d) Use Gancu

Gancu attributes that can be categorized as follows:

1) There is Gancu: when transferring fertilizer from the car to the warehouse or from the warehouse to the car, the fertilizer in the sack is damaged so that it often leaks or spills.
2) No Gancu: the transfer of fertilizer is carried out from the car to the storage warehouse or from the warehouse to the car again, the fertilizer sack/burlap is safe or there is no leakage or spilled fertilizer.

The attributes above are the result of filtering that has been done so that four important attributes are needed for the decision on the feasibility of shrinking the volume of fertilizer in the storage warehouse. This filtering activity includes removing incomplete data, removing unnecessary attributes and removing duplicate data. The next activity is to

combine data for eligible and unfeasible categories.The next analysis of data needs, namely the form of output that will be produced can be seen in the table below:

Table 2. Final Result of Feasibility of Fertilizer Volume Depreciation

No	Decision
1	Exist
2	There is not any

The results of the filtering activities can be seen in the following table:

Table 3. Sample Data From Filtering Results

Place to Save (Location)	Save Time	Fertilizer Composition	Use Gancu	Decision
Moist	1	SB	PG	Exist
Moist	1	SB	TAG	There is not any
Moist	1	SP	PG	Exist
Moist	1	SP	TAG	There is not any
Moist	2	SB	PG	Exist
Moist	2	SB	TAG	There is not any
Moist	2	SP	PG	Exist
Moist	2	SP	TAG	There is not any
Moist	3	SB	PG	Exist
Moist	3	SB	TAG	Exist
Moist	3	SP	PG	Exist
Moist	3	SP	TAG	Exist
Dry	1	SB	PG	Exist
Dry	1	SB	TAG	There is not any
Dry	1	SP	PG	Exist
Dry	1	SP	TAG	There is not any
Dry	2	SB	PG	Exist
Dry	2	SB	TAG	There is not any
Dry	2	SP	PG	Exist
Dry	2	SP	TAG	There is not any
Dry	3	SB	PG	Exist
Dry	3	SB	TAG	Exist
Dry	3	SP	PG	Exist
Dry	3	SP	TAG	Exist
hot	1	SB	PG	There is not any
hot	1	SB	TAG	There is not any
hot	1	SP	PG	There is not any
hot	1	SP	TAG	There is not any
hot	2	SB	PG	Exist
hot	2	SB	TAG	There is not any
hot	2	SP	PG	Exist
hot	2	SP	TAG	There is not any
hot	3	SB	PG	Exist
hot	3	SB	TAG	Exist
hot	3	SP	PG	Exist
hot	3	SP	TAG	Exist

After the data is in the filtering above, then processing is carried out by grouping the data as follows:

1) Grouping the value of storage area, grouping the value of this storage area based on the location or place of storage of fertilizer in the warehouse, so that the value can be grouped as in table 4. below:

Table 4. Classification of Depreciation Time Value

Storage	Classification
Humid Location	L
Dry Location	K
Hot Location	P

2) Grouping storage time values, grouping storage time values based on storage time so that these values are grouped as in table 5. below:

Table 5. Classification of Storage Values

Storage Time	New Value
Day	1
Sunday	2
Month	3

3) Grouping the value of the composition of fertilizers, grouping the values of the composition of fertilizers based on the composition of fertilizers so that these values can be grouped as in table 6 below:

Table 6. Classification of Fertilizer Composition Values

Fertilizer Composition	New Value
Regular Arrangement	SB
Procedure Arrangement	SP

4) Grouping the value in use of gancu, grouping the value in use of gancu is based on the gancu used so that the value can be grouped as in table 7. below:

Table 7. Classification of the Use Value of Gancu

Use Gancu	New Value
There's a buckle	AG
No Buckles	TAG

4. Conclusion

From the descriptions in the chapters already discussed earlier it can be concluded:

a) Algorithm C4.5 with decision tree method can provide executive information and the system is used to describe the process related to the classification of fertilizer volume depreciation in the storage warehouse of PT. This Young Iskandar.

b) Algorithm C4.5 is considered as an algorithm that is very helpful in classifying data because the characteristics of the classified data can be obtained clearly, either in the form of decision tree structure (decision tree) or rules or rule-if-then, making it easier for users to perform information mining on the relevant data

c) The form of the decision tree in providing recommendations on decision - making for the feasibility of reducing the volume of fertilizer in the storage warehouse.
d) The system developed can help in the classification to determine any factors that cause a decrease in the volume of fertilizer in the storage warehouse so that the management of PT. Iskandar Muda fertilizer can take decisions that are the complaints of farmers in the purchase of fertilizer.

References

[1] Agrawal Vineet S, Rajesh M, Sonali K, Mukesh P. A contemporary overview of endodontic irrigants – A review. Journal of Dental application 2014; 1(6): 105- 15.
[2] Aravind, G., Bhowmik, D., Duraivel, S., & Harish, G., 2013, Traditional and Medicinal Uses of Carica papaya, Journal of Medicinal Plants Studies, 1 (1), 7-15.
[3] Berry, Michael J.A dan Linoff, Gordon S. 2004. Data Mining Techniques For Marketing, Sales, Customer Relationship Management Second Editon. United States of America: Wiley Publishing, Inc.
[4] Han & Kember : 2006 Data Mining Concept and Technigues, 2nd ed.
[5] Kusrini, luthfi taufiq Emha, (2009), Algoritma Data Mining, Penerbit Andi, Yogyakarta.
[6] Larose D, T., 2005, Discovering knowledge in data : an introduction to data mining, Jhon Wiley & Sons Inc.
[7] Turban. E, dkk. 2005, "Decision Support System and Intelligent Systems", 7th Edition, Jilid 1, Yogyakarta: Andi.

Author

Yendrizal, S.Kom., M.Kom, Born on November 27, 1990 in Ranggomalai Agam. Completed SI and S2 studies at the University of Putera Indonesia. Started teaching at AMIK KOSGORO on August 6, 2007 until now. Courses that have been in powerful Visual Basic Programming, Computer Installation and Graphic Design.