**IJISCS (International Journal of Information System and Computer Science)**
IJISCS, Volume 7, Number 2, August 2023
E ISSN: 2598-246X; P ISSN: 2598-0793, pp. 161-167
**Accredited SINTA 4 Number 79/E/KPT/2023**
https://jurnal.ftikomibn.ac.id/index.php/ijiscs

# THE C45 ALGORITHM METHOD IN PREDICTING DAMAGED GOODS CASE STUDY: SEMULI RAYA INDOMARET SHOP

**Dessry Maeye Khelany[1*], Nurmayanti[2], Sigit Mintoro[3]**
[1]Information System, ITBA Dian Cipta Cendikia Kotabumi
[2,3]Technology Computer, ITBA Dian Cipta Cendikia Kotabumi
[1,2,3]Lintas Sumatera Street, No.03, Candimas, Kotabumi, North Lampung

*Corresponding author
dessrymaeyekhelany@gmail.com[1*]
nurmayanti@gmail.com[2]

**Article history**:
Received June 12, 2023
Revised July 27, 2023
Accepted August 31, 2023

**Keywords**:
Damaged Goods;
Data Mining;
Algorithm C4.5;
Prediction.

**Abstract**
Indomaret Semuli Raya is a company that competes in the industrial world. In the industrial world, the quality of products marketed is an important indicator for Indomaret Semuli Raya to be able to stand during intense competition from other companies. Product quality is certainly the thing that attracts consumers. In dealing with problems that occur in the company, it must make the right decision in determining the product strategy to be sold, to get the right decision, sufficient item data is needed to be analyzed. This research raises the issue of whether or not goods are worth selling in the category of damaged goods at Indomaret Semuli Raya in the period from February to March 2023. Data mining is used in this study, specifically the C4.5 Algorithm Method. The author chose the C4.5 Algorithm method because it can be used to determine whether or not damaged goods are unfit for sale. Algorithm C4.5 will determine damaged goods based on the attributes that become the standard for the eligibility of goods. The attributes referred to in this study are Overweight, Brackage, Sell by, and Moisture & Temperature. The aftereffects of manual computations by means of Microsoft Succeed utilizing the C4.5 Calculation have a precision of 90.00% then demonstrated by the RapidMiner application with an exactness of 90.00%.

## 1.0 INTRODUCTION

Product quality is an important indicator for a company to be able to stand during intense competition in the industrial world. Product quality is solely determined by consumers so consumer satisfaction can only be achieved by providing good quality. The quality of a product is built by the company by taking into account the needs and desires of the customer because an industrial factory will not exist if the products made or ordered are not following the wishes of consumers[1].

In dealing with problems that occur in the company, it must make the right decision in determining the product strategy to be sold, to get the right decision, sufficient item data is needed to be analyzed. Data mining is the process of mining or finding new information from large amounts of data by looking for certain patterns or rules that are expected to overcome these conditions. Classification is one of the many methods in data mining itself. The Decision Tree is a component of the classification method, which is comprised of several methods. There are various methods for doing Data Mining that are used such as Naïve Bayes and C.45. The strategy contained in the information mining that will be utilized in this study is the C4.5 calculation. the process by which a decision tree (Decision Tree) is constructed. Researchers

attempt to resolve existing issues by utilizing the C4.5 algorithm and classification techniques based on the aforementioned issues[2].

The C.45 algorithm, also known as the Decision Tree or C4.5 algorithm, is one of the most widely used classification techniques due to its ease of human intervention. A choice tree is a prescient model utilizing a tree structure or various leveled structure. Data can be transformed into decision trees and decision rules using the idea of a decision tree. A decision tree typically presents the data in the form of attributes and records in a tabular format. When forming a tree, attributes specify a parameter that is used as a criterion[3].

From the problems described above, the authors conducted research with the title "Implementation of the C4.5 Algorithm Method in Predicting Damaged Goods Case Studies; Toko Indomaret Semuli Raya" incorporates data mining and the C4.5 classification algorithm into a system that will later serve as a decision support tool for Indomaret Semuli Raya's classification of damaged goods.

## 2.0 THEORETICAL

One of the most important aspects of data mining is classification. A set of training data with a predetermined class is used to create a classifier. The most common way of tracking down models (or works) that portray and separate information classes or ideas that intend to be utilized to foresee classes of items whose class marks are obscure. Decision/classification trees, Bayesian/Naive Bayes classifiers, neural networks, statistical analysis, genetic algorithms, rough sets, k-nearest neighbors, rule-based methods, memory-based reasoning, and Support vector machines (SVM) are examples of widely used classification algorithms[3].

To maintain the quality of goods and stock availability, preventive measures are needed to reduce the level of damage caused by the high mobility of goods during the delivery process. The aspects that need to be considered here are prevention during the process of handling goods while in the storage warehouse, the tools used, and the awareness of each employee about the quality of the goods to be sent. Because every consumer will expect a quality product with good conditions[4].

Preprocessing Data is the initial process of data processing. In data mining, the quality of the data to be used needs to be considered. There are several data preprocessing factors that affect data quality, including accuracy, integrity, consistency, actuality, and interpretation Data preprocessing is a basic operation for completing consistent data without being noisy[5].

Data mining is the process of extracting and identifying useful information and related knowledge from large databases using statistics, mathematics, AI, and machine learning[6]. Algorithm C4.5 is an algorithm that builds decision trees and forms knowledge models to classify data and Algorithm C4.5 has the fastest performance and has the highest accuracy. Data Mining, Classification, Naive Bayes, Chi squared In light of the consequences of the examination performed, rapid-miner programming can help and make it more straightforward to deliver probabilities to be utilized as expectations[7].

### 2.2. Theories system used

A unified environment for machine learning, deep learning, text mining, and predictive analytics is provided by Rapid Miner, a data science software platform developed by the same name company. It supports all steps of the machine learning process, including data preparation, results visualization, validation, and optimization, and it is utilized for business and commercial applications as well as research, education, training, rapid prototyping, and application development. RapidMiner was created with an open center model[7].

## 3.0 METHODOLOGY

### 3.1 Research Design

In this review, demonstrating will be completed involving the C4.5 Calculation technique for handling thing information, then for dataset handling, and for handling the consequences of the precision of the calculation utilized. Calculation technique C4.5 was picked in light of the fact that one of its benefits is that it can deal with numeric and discrete information. Gain ratio is used in algorithm C4.5. Prior to computing the increase proportion, it is important to work out the data esteem in bits from an assortment of articles, specifically by utilizing the idea of entropy to shape a choice tree. Then, an algorithm is used to calculate the data in

accordance with the method, and accurate results are sought. In this study, the stages of the modeling process are as follows:

1. Choosing the fitting Information Mining task
   At this point, the author selects the type of data mining that will be used for classification and the method that will be used to identify damaged (expired) goods.
2. Choosing the Data Mining Algorithm
   After selecting the type of data mining to be used, namely classification, then the next step is to determine the classification algorithm to be used. In this study, the algorithm chosen for classification is Algorithm C4.5.
3. Employing the Data Mining Algorithm
   This stage is carried out for data processing with the algorithm that has been selected, namely by using the C4.5 Algorithm.
4. Evaluation
   Patterns derived from the results of the algorithms used to determine rules, reliability, and other factors are evaluated and interpreted at this stage. Assessment is done by applying the example got from the past cycle to the testing information gave. Assessment is finished by utilizing a disarray grid and ROC bend.
5. Discovered Knowledge
   Patterns derived from the results of the algorithms used to determine rules, reliability, and other factors are evaluated and interpreted at this stage. Assessment is done by applying the example got from the past cycle to the testing information gave. Assessment is finished by utilizing a disarray grid and ROC bend.

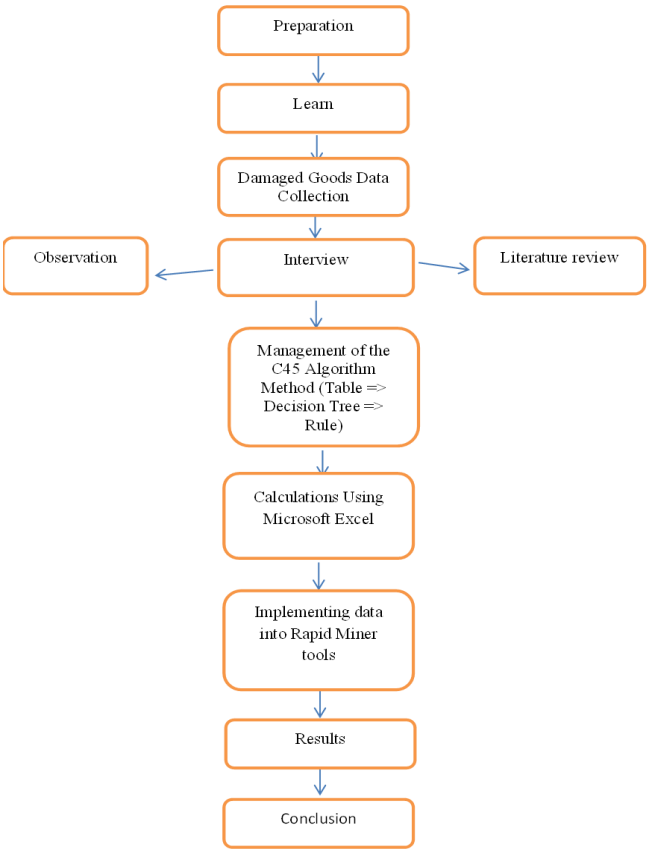


**Figure 1**. Frame of Mind

The C4.5 Algorithm is used in the proposed model for determining goods damage at Indomaret Semuli Raya. Algorithm C4.5 will typically construct a decision tree. Prepare training data. The training data is used to form the model to be applied to the test data.
Determine the roots of the tree.

$$\text{Entropy (S)} = \sum_{i=1}^{n} - pi * \log_2 pi$$

### 3.2 Rapid Miner

A unified environment for machine learning, deep learning, text mining, and predictive analytics is provided by Rapid Miner, a data science software platform developed by the same name company. It aids in the preparation of data, visualization of results, validation, and optimization, as well as in research, education, training, rapid prototyping, and application development, and it is utilized in both business and commercial applications. An open-core model was used in the development of RapidMiner[9].

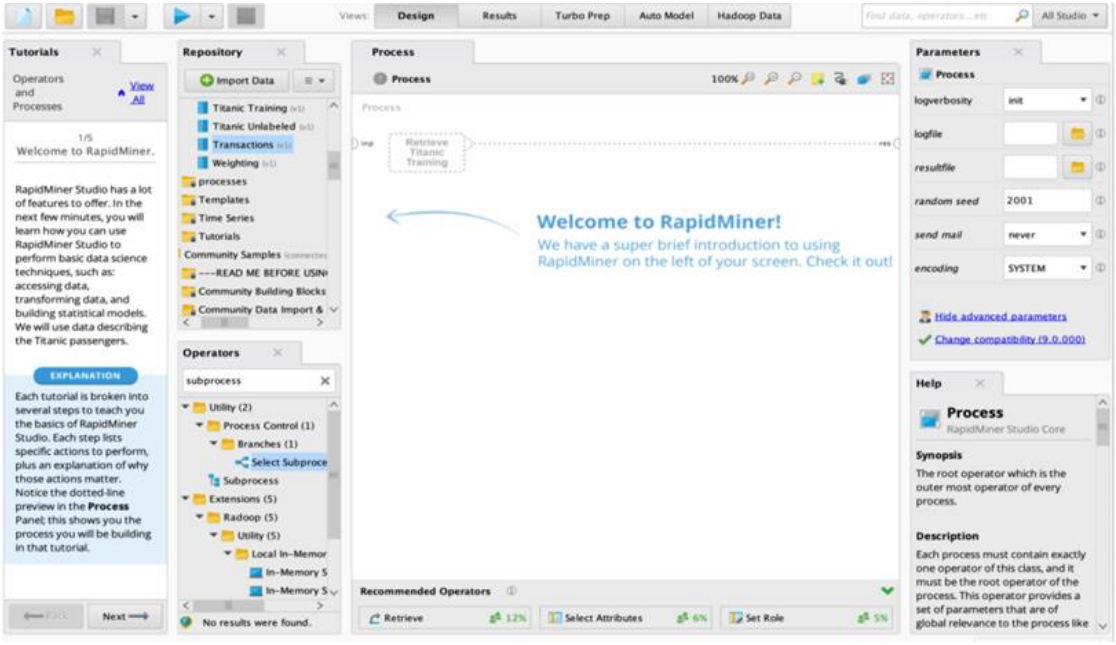


**Figure 2** Rapid-miner

## 4.0 RESULTS

### 4.1 Enthropy and Gain Result

The following is the prediction calculation with 4 attributes, namely: Overweight, Breakage, Sell By, and Moisture & Tempetarure Using Microsoft Excel 2021. How to implement the c4.5 algorithm with Microsoft Excel 2021 is as follows[10].

1. Open the Microsoft Excel 2021 application on a PC/Laptop
2. Next, on the worksheet, make a data table for damaged goods based on the attributes that have been determined. The attributes in this study are Overweight, Brake, Sell By, Moisture & Temperature.
3. Next, enter the list of goods in the table and calculate the number of temporary labels.
4. After calculating the total amount, then make the Entrophy and Gain tables.
5. To find Entrophy in Excel, enter the formula in the formula column, namely: =((SUMUM YES /ALL TOTAL)*IMLOG2(SUMUM YES / ALL TOTAL)+(-SUMUM NO / ALL TOTAL)*IMLOG2(SUMUM NO / ALL TOTAL) )
6. After the Entrophy values of all attributes are completed, then look for the Gain value
7. To find the following Gain value, an example of the overweight attribute, enter the formula in the formula column, namely: =(ENTROPHY TOTAL) – ((NUMBER OF NEAT / ALL TOTAL) * ENTROPHY NEAT) – ((NUMBER OF DENT / ALL TOTAL) * ENTROPHY DENT ) – (( TOTAL TEAR / ALL TOTAL) * ENTROPHY TEAR)
8. 8. Next, enter the entropy and gain formulas for each attribute.
9. 9. After the Entrophy and Gain values are complete, then look for the probability value by entering the formula.

**Table 1** Entropy and Gain

| Atribute | | Count | Yes | No | *Entrophy* | *Gain* |
|---|---|---|---|---|---|---|
| **TOTAL** | | 89 | 26 | 63 | 0,871463006 | |
| ***Overweight*** | | | | | | 0,27441484 |
| | Neat | 47 | 22 | 22 | 0,997059057 | |
| | Dent | 29 | 1 | 28 | 0,216396932 | |
| | Tear | 13 | 0 | 13 | 0 | |
| ***Breakage*** | | | | | | 0,287193343 |
| | Normal | 52 | 26 | 26 | 1 | |
| | Leaky | 16 | 0 | 16 | 0 | |
| | Deaflated | 21 | 1 | 21 | 0 | |
| ***Sell By*** | | | | | | 0,313860104 |
| | Before | 49 | 26 | 20 | 1,012788944 | |
| | Expierd | 40 | 0 | 40 | 0 | |
| **Moisture & Temperature** | | | | | | 0,347692572 |
| | Safe | 47 | 26 | 21 | 0,991820797 | |
| | Smell | 21 | 0 | 21 | 0 | |
| | Fade | 21 | 0 | 21 | 0 | |
| **Gain Max** | | | | | | 0,347692472 |

From the table above there are 89 data consisting gain value located in moisture and terperature with a gain value of 0.34769 so the most prominent value is moisture and temperature. This shows that calculations using excel are far more accurate than manual ones.

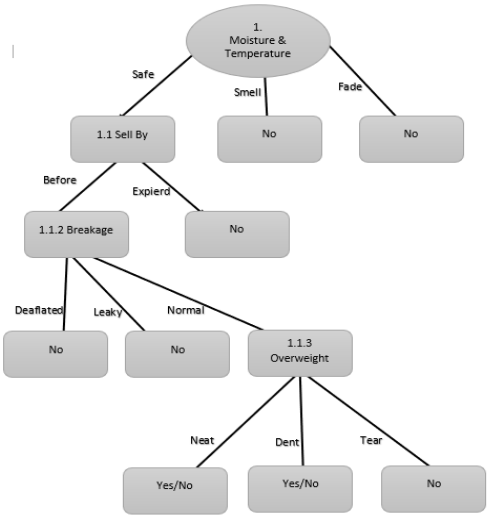


**Figure 3** Decision Tree

The decision tree has node 1 on moisture and temperature where 1.1 sell by is divided into 1.1.2 breakege which is further divided into 1.1.3 overweight, 3 aspects of neast dent tears emerge. Yes no yes no.

**Table 2** Confusion Matrix

| | Prediction | |
|---|---|---|
| Actual | Yes | No |
| Yes | **0** | **0** |
| No | **9** | **1** |
| Accuracy | **90,00%** | |

Rapid-miner calculation steps:

The picture above is the formula used in the Rapid Minner 10.1.2 application to implement the C4.5 Algorithm. Select Retrive, drag and drop to the worksheet, then rename it with training and testing data in the Operator column as the first step. Next, enter the data that has been input into the rapid-miner on the retrieval operator. Next, select the Set Role operator then drag and drop it to the worksheet. In training data, connect it to exa on set roll. Then select the

Decision Tree on the operator then drag and drop on the worksheet, connect the exa to the tra Decision tree. Then select Apply model on the drag and drop operator to the worksheet then connect the mood in the decision tree to the mood apply model mood. Next, also connect the testing data to the apply model. Select the crossnValidation operator, connect set role 2 exa to exa Cross Validation. Next connect mood, exa, test, per to res. Next play[12].

1. After carrying out the above steps, the results of the prediction will appear as follows.

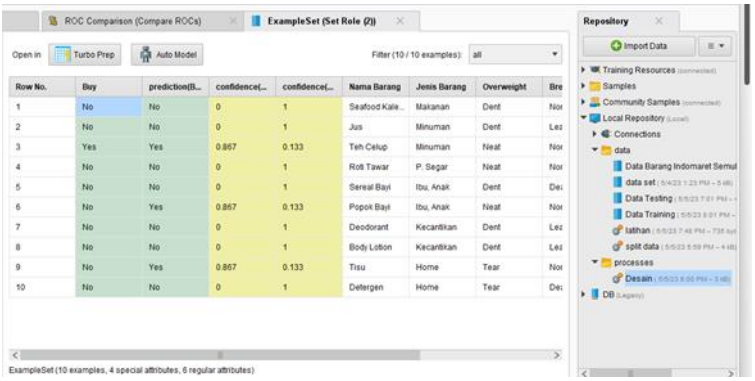


**Figure 4** Data Testing

Then connect each ports then after all of them are connected then we can test it by clicking the start or run button. From the previous process results appear as shown above, namely the prediction table[13][14][15].

2. Accuracy results of RapidMiner calculations with Accuracy of 90.00%

**accuracy: 90.00% +/- 31.62% (micro average: 90.00%)**

| | true No | true Yes | class precision |
|---|---|---|---|
| pred. No | 9 | 1 | 90.00% |
| pred. Yes | 0 | 0 | 0.00% |
| class recall | 100.00% | 0.00% | |

**Figure 5** Accuracy

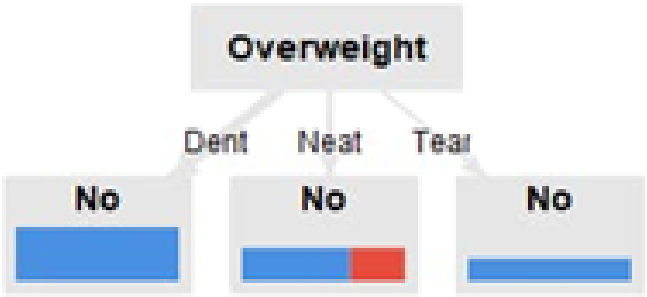


**Figure 6**. Decision Tree

## 5.0 CONCLUSION

Based on the discussion in the previous chapters, the accompanying ends can be drawn. At Indomaret Semuli Raya, the Algorithm C4.5 Classification Method can be used to predict the accuracy of the Damaged Goods Data. In this study the classification used was Microsoft Excel 2021 and Rapid Minner from calculations carried out in Microsoft Excel 2021 using the C4.5 Algorithm has an accuracy of 90.00% And then the RapidMiner application is proven with the Retive, Set role operator, Apply Model, Decision Tree and Cross Validation with results of 90.00% accuracy as well. The accuracy results of manual calculations and RapidMiner calculations show the same final result. This means that the research on the classification of damaged goods at Indomaret Semuli Raya uses Microsoft Excel and Rapid Minner in

implementing the C4.5 algorithm, the results are accurate. Decision tree Algorithm C4.5 uses RapidMiner

## REFERENCES


[1] M. T. Informatika and U. Amikom, "ANALISIS PEMBOBOTAN KATA PADA KLASIFIKASI TEXT MINING," vol. 3, no. 2, pp. 179–184, 2019.

[2] I. Technology and C. Science, "No Title," vol. 4, pp. 20–29, 2021.

[3] H. Annur, "KLASIFIKASI MASYARAKAT MISKIN MENGGUNAKAN METODE," vol. 10, pp. 160–165, 2018.

[4] I. A. Nikmatun, U. Diponegoro, I. Waspada, and U. Diponegoro, "IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," vol. 10, no. 2, pp. 421–432, 2019.

[5] A. F. Cahyanti, "Penentuan Model Terbaik pada Metode Naive Bayes Classifier dalam Menentukan Status Gizi Balita dengan Mempertimbangkan Independensi Parameter," vol. 4, no. 1, pp. 28–35, 2015.

[6] R. Yusri, S. Edriati, and R. Yuhendri, "Rangkiang : Jurnal Pengabdian Pada Masyarakat UP3M STKIP PGRI Sumatera Barat Rangkiang : Jurnal Pengabdian Pada Masyarakat UP3M STKIP PGRI Sumatera Barat," vol. 2, no. 1, pp. 32–37, 2020.

[7] N. Zumel and J. Mount, "Practical Data Science with R."

[8] N. Pati, "ALGORITMA C4 . 5 UNTUK PENJURUSAN SISWA SMA," pp. 3–6.

[9] S. V. O. L. I. No, O. Algoritma, C. Klasifikasi, and P. P. Jantung, "No Title," vol. I, no. 1, pp. 26–36, 2014.

[10] A. I. Waspah *et al.*, "EXPECTATION MAXIMIZATION ALGORITHM MEMPREDIKSI PENJUALAN SUSU MURNI PADA PT . SEWU PRIMATAMA INDONESIA LAMPUNG," vol. 7, no. 1, pp. 27–38, 2022.

[11] C. Algorithm *et al.*, "Classification and Clustering of Internet Quota Sales Data Using," vol. 9, no. 2, pp. 268–283, 2023, doi: 10.26555/jiteki.v9i2.25970.

[12] E. V. Astuti and R. Mawarni, "THE COMPARISON USING EXPECTATION- MAXIMIZATION ALGORITHM AND C4 . 5 ALGORITHM TO PREDICT THE RESULT OF BIOGAS PRODUCTION AS A POWER PLANT AT PT BUDI STARCH & SWEETENER ( BSSW )," pp. 186–198.

[13] A. P. Wibawa, M. Guntur, A. Purnama, M. F. Akbar, and F. A. Dwiyanto, "Metode-metode Klasifikasi," vol. 3, no. 1, pp. 134–138, 2018.

[14] J. Media and I. Budidarma, "Data Mining Menggunakan Algoritma K-Nearest Neighbor Dalam Menentukan Kredit Macet Barang Elektronik," vol. 5, pp. 1063–1067, 2021, doi: 10.30865/mib.v5i3.3100.

[15] D. S. Seruni, M. T. Furqon, and R. C. Wihandika, "Sistem Prediksi Pertumbuhan Jumlah Penduduk Kota Malang menggunakan Metode K-Nearest Neighbor Regression," vol. 4, no. 4, pp. 1075–1082, 2020.