C4.5 ALGORITHM IN PREDICTING THE SUCCESS RATE OF FIG CULTIVATION

Rika Nofitri[1], Novica Irawati[2]

[1]Computer System, Sekolah Tinggi Manajemen Informatika dan Komputer Royal, Indonesia
[2]Information System, Sekolah Tinggi Manajemen Informatika dan Komputer Royal, Indonesia

Corresponding author:
nofitririka15@gmail.com

ABSTRACT

Figs (Ficus carica L.) is a type of edible fruit-producing plant originating from West Asia. Fig is a rare plant with rich benefits that has a high enough price. The main problem often faced by fig farmers is the difficulty in determining a suitable formulation between planting media, watering, sunlight and temperature. This causes many failures and disadvantages fig farmers. This study aims to determine the best formula for fig plants in order to reduce the risk of failure and obtain good quality fig. Because many experiments have been carried out by fig farmers with varying percentages of success, then to facilitate farmers, especially beginner farmers, in making a decision then the C4.5 data mining algorithm is used to help produce the right formula in fig cultivation. Algortima c4.5 will produce a decision tree that contains rules about what criteria play an important role in the success of the cultivation.

INTRODUCTION

The development of technology, especially in the field of computers today makes it easy for us to do various activities, ranging from easy things to complicated ones. One of them is in agriculture. Tin or Ara (Ficus carica L.) is a type of edible fruit-producing plant originating from West Asia. Tin is a rare plant with rich benefits that has a high enough price.[1]

The main problem often faced by tin farmers is the difficulty in determining a suitable formulation between planting media, watering, sunlight and temperature. This causes many failures and disadvantages tin farmers. Based on the description above, the researcher is interested in helping to solve the problem through the process of extracting the results of experimental data and experiences that have been carried out by tin farmers. Data mining will be performed using data mining algorithm C4.5. through this method a decision tree will be obtained which contains rules about what criteria play an important role in the success of the planting [2].

METHOD

Data Mining is a process of finding useful patterns or knowledge automatically from a large collection of data.[3] C4.5 algorithm is the development of ID3. Because of this development the C4.5 algorithm has the same basic principles of work as the ID3 algorithm. It's just that in the C4.5 algorithm the attribute selection is done using the Gain Ratio.[4]

In general the C4.5 algorithm for building decision trees is as follows:
a. Select the attribute as the root
b. Create a branch for each value
c. Divide cases in branches
d. Repeat the process for each branch until all cases in the branch have the same class.

To choose an attribute as the root, based on the highest gain value of the existing attributes. To calculate the gain the following formula is used:

$$Gain\,(S, A) = Entropy\,(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si) \qquad (1)$$

Where:
S = space (data) sample used for training
A = attribute
N = number of partitions A
| Si | = number of cases in the first case i
| S | = number of cases in S

While the entrophy value calculation can be seen in the following formula:

$$Entropy\,(S) = \sum_{i=1}^{n} -pi * \log 2\, pi \quad (2)$$

Where :
S = case set
n = number of partitions S
pi = proportion of Si to S

Rapid miner is a data science software platform developed by the company of the same name, which provides an integrated environment for machine learning, deep learning learning), text mining (text mining), and predictive analytics. This application is used for business and commercial applications as well as for research, education, training, manufacturing prototype quickly, and application development and support all steps of the machine learning process including data preparation, results visualization, validation and optimization. Rapid Miner was developed with an open source model. [5] The following is an overview of the input and target variables:
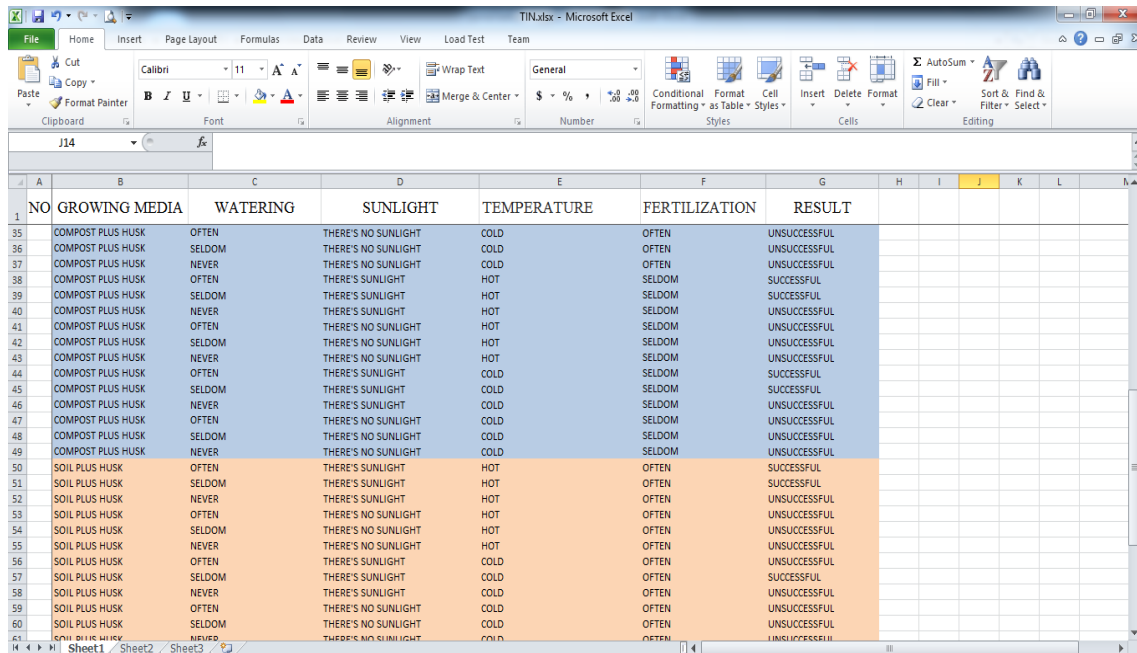
Table 1. Input Variables and Target Variables

INPUT VARIABLES					TARGET VARIABLE
GROWING MEDIA	WATERING	SUNLIGHT	TEMPE RATURE	FERTI LIZATION	RESULT
SOIL	OFTEN	THERE'S SUNLIGHT	HOT	OFTEN	UNSUCCESSFUL
SOIL	SELDOM	THERE'S SUNLIGHT	HOT	OFTEN	SUCCESSFUL
COMPOST PLUS HUSK	OFTEN	THERE'S SUNLIGHT	HOT	OFTEN	SUCCESSFUL
COMPOST PLUS HUSK	SELDOM	THERE'S SUNLIGHT	HOT	OFTEN	SUCCESSFUL
SOIL PLUS HUSK	OFTEN	THERE'S SUNLIGHT	HOT	OFTEN	SUCCESSFUL
SOIL PLUS HUSK	OFTEN	THERE'S NO SUNLIGHT	HOT	OFTEN	UNSUCCESSFUL

The first step is to create a database that contains data for input variables and target variables in .xls format. After that, import the data into the software, then connect the data to the decission tree, then execute it, a decision tree will be formed.

RESULT AND DISCUSSION
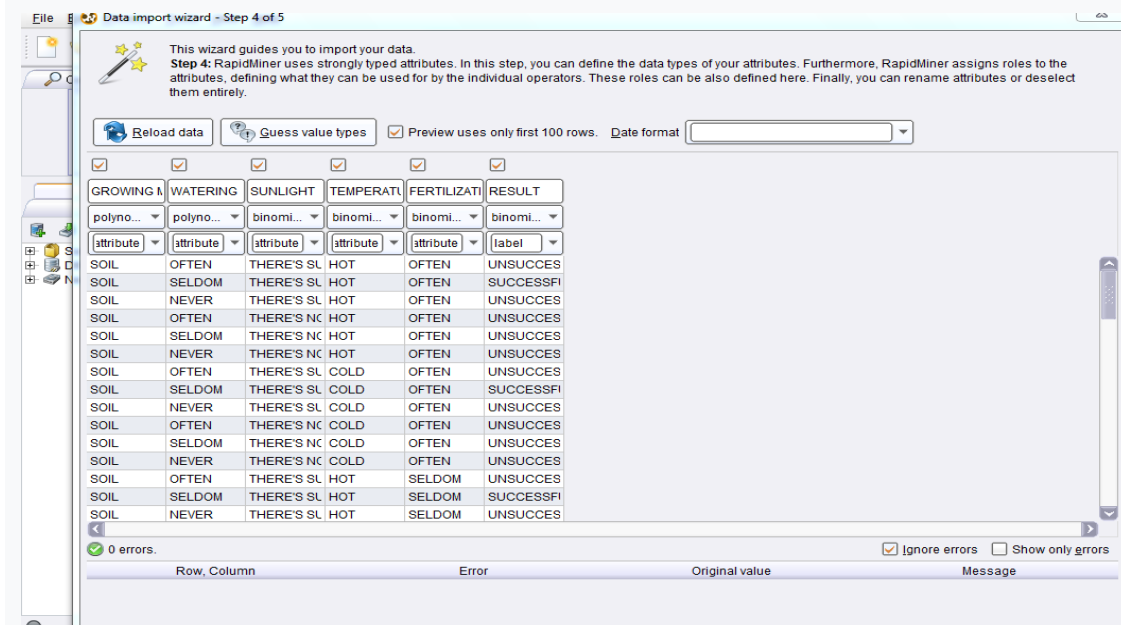
Based on the results of data collection that has been done, to meet the needs of the system, the data transformation is carried out in advance decision system. This decission system will be imported into tools that have been determined by researchers, namely using rapidminer software. As for the imported data which is intended as follows:

Image 1. Excel Data

The picture above is a display of data stored in Microsoft Excel. This data will be imported into Rapidminer to form a decision tree.

Image 2. Data Import Wizard

In Figure 2 above, you can see the condition attributes of the results of the profits, namely the turnover, purchase, stock, and costs. While the decision attribute is the result of cultivation which is marked by a label mark on the wizard's form. After that

in the application of rapidminer software that is the process of connecting between imported data with the operator decision tree method in the operator menu. Then drag the data and operator or make the process of moving into the process window. Then the connection can be seen in the picture as follows.
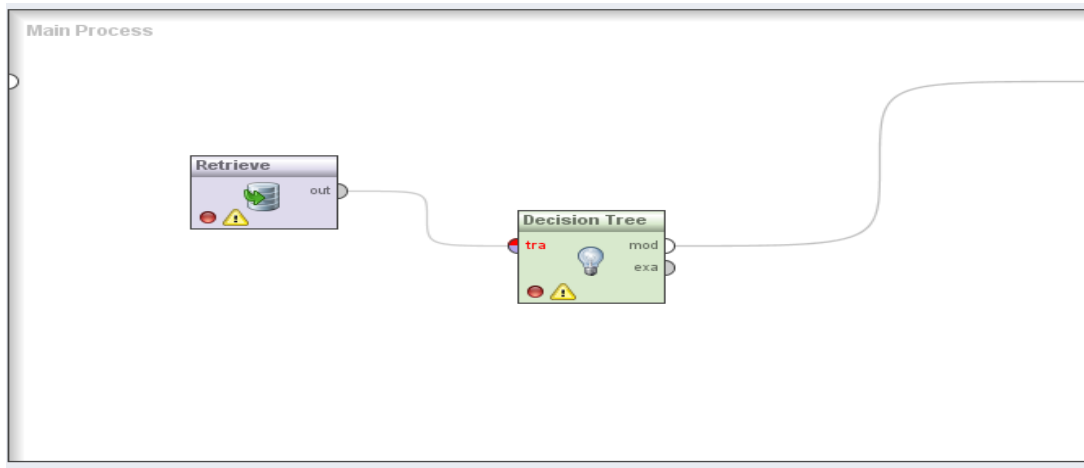

Image 3. Import Data Connection with C4.5 Algorithm

From the steps shown in Figure 3, the next process is to run the application by clicking the run button on the application window. So from the results of the process carried out the results obtained are decision trees that describe the relationship between the criteria that lead to the success or failure of figs cultivation.
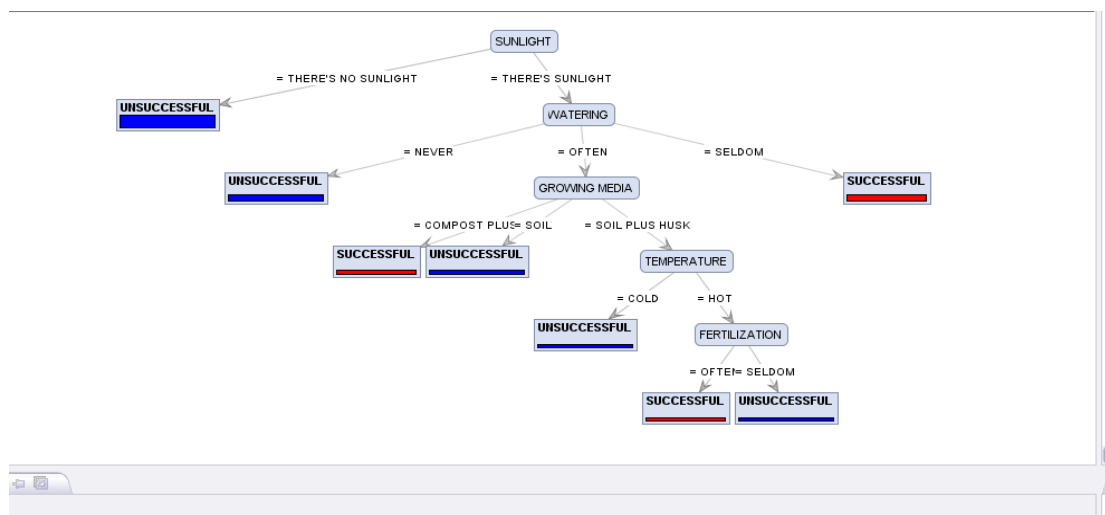

Image 4. Decission Tree

Image 4 above is a picture of a decision tree formed. From the picture above we get the following rules:
1. IF there isn't sunlight THEN the result is unsuccessful
2. IF there is sunlight AND watering is seldom THEN the result is successful

3. IF there is sunlight AND watering is never THEN the result is unsuccessful
4. IF there is sunlight AND watering is often AND growing media are compost plus husk THEN the result is successful
5. IF there is sunlight AND watering is often AND growing media is soil THEN the result is successful
6. IF there is sunlight AND watering is often AND growing media are soil plus husk AND temperature is cold THEN the result is unsuccessful
7. IF there is sunlight AND watering is often AND growing media are soil plus husk AND temperature is hot AND fertilization is seldom THEN the result is unsuccessful
8. IF there is sunlight AND watering is often AND growing media are soil plus husk AND temperature is hot AND fertilization is often THEN the result is successful

CONCLUSION

The decision tree in providing recommendations for decision making for the success of fig cultivation is seen as the first criteria that becomes the root, namely sunlight. The criteria for sunlight have a second sub root, namely watering. Then the subroot of watering is the planting medium. Then the subroot of the planting medium is temperature. Furthermore, the last sub root is fertilization

BIBLIOGRAPHY

[1]   A. Aziz, "Penerapan algoritma c4.5 pada klasifikasi hama dan penyakit tanaman bawang merah," *J. Tek.*, vol. 10, no. 2, 2003.

[2]   N. Khasanah, "KANDUNGAN BUAH-BUAHAN DALAM ALQUR'AN: BUAH TIN (Ficus carica L), ZAITUN (Olea europea L), DELIMA (Punica granatum L), ANGGUR (Vitis vinivera L), DAN KURMA (Phoenix dactylifera L) UNTUK KESEHATAN," *Phenom. J. Pendidik. MIPA*, vol. 1, no. 1, p. 5, 2016, doi: 10.21580/phen.2011.1.1.442.

[3]   M. A. Sembiring and Z. Azhar, "Factors Analysis And Profit Achievement For Trading Company By Using Rough Set Method," *Int. J. Artif. Intell. Res.*, vol. 1, no. 1, p. 16, 2017, doi: 10.29099/ijair.v1i1.15.

[4]   S. M. S. Sianturi and N. A. Hasibuan, "ANALISA DATA PERTANIAN TANAMAN PANGAN UNTUK MEMPREDIKSI HASIL PANEN DENGAN DATA MINING PANGAN dan HOLTIKUTURA PROVINSI SUMUT)," *J. Pelita Inform.*, vol. 18, no. April, pp. 212–219, 2019.

[5]   E. Darmawan, "C4 . 5 Algorithm Application for Prediction of Self Candidate New Students in Higher Education," vol. 3, no. 1, pp. 22–28, 2018, doi: 10.15575/join.v3i1.171.