Comparison of Case-Based Reasoning and Dempster Shafer on Expert System of Cassava Disease Identification

Minarni*, Indra Warman, Yuhendra and Wenda Handayani

Institut Teknologi Padang, Informatics Engineering Study Program, Padang, Indonesia

Abstract. This study aims to develop a web-based expert system that can identify cassava disease using Case-Based Reasoning Method and Shafer Dempster Method, and to know the performance comparison of both based on the accuracy of identification. Case-Based Reasoning (CBR) is a computer-generated system that uses old knowledge to solve new problems. CBR provides solutions to new cases by looking at the oldest cases that are closest to new cases. The identification process is done by entering new cases containing the symptoms to be identified into the system, then perform the process of calculating the value of similarity between the new case and the base case using the nearest neighbor method. Dempster Shafer based on two ideas is the idea of obtaining degrees of confidence of subjective possibilities and the rule of dempster safer itself to combine degrees of confidence based on the evidence obtained. This expert system is built using PHP programming language and MySQL data base. The output of the system is the percentage of identification result of both methods. Testing and analysis results show that Case-Based Reasoning provides better identification accuracy than Dempster Shafer.

1 Introduction

Cassava (Manihot Esculenta Crantz) is one of the plants that can be consumed tubers and leaves. Cassava can be used as basic material of food and processed raw materials of home industry. One of the utilizations of cassava as the main ingredient of typical food of West Sumatra. Therefore, there is an effort to make cassava productivity not decrease. Based on data from Food Security Agency (Badan Ketahanan Pangan) of West Sumatra, cassava has decreased production by 19.22%. One of the causes of declining Indonesian cassava production is pests and diseases such as leaf mites, fungi and bacteria. Pest and disease attacks are caused by the wrong cropping pattern and the lack of knowledge of the farmers will attack pests and diseases. The lack of knowledge of the farmers, especially in terms of identifying pests and diseases in cassava.

To identify pests and diseases of cassava plants can be known from the symptoms that appear. The characteristics of plants affected by disease with other diseases are so similar that it confuses people to identify them. Conversely there are also plants affected by diseases with different characteristics but still confusing in remembering the name and handling of the disease. Handling of pests and diseases is done by experts in this field commonly known as Agricultural Extension Field (PPL). Data from the Government Agency's Performance Accountability Report (LAKIP) of the Food Crops Office of West Sumatera Province in 2015 that the number of Agricultural Field Agricultural Extension (PPL) for food crops and horticulture is only 234 people, while the number of farmers is 644,610 people [1]. From the data, it appears that the number of extension workers with farmers is not comparable. In addition, trainers are constrained by time and place, so a system is needed that can help farmers to identify, pests and diseases of cassava plants.

The expert system has the ability to facilitate practical problems when the expert is absent. And one of the implementation of expert systems in agriculture is to identify plant diseases. Research on expert system of plant disease diagnosis has been done by many previous researchers, including: Abu-Naser (2008) developed an expert system using two methods of diagnosis of plant diseases, namely descriptive method and graphics. Descriptive method is based on the symptoms experienced by the plant. While on the graphical method based on the picture. The result obtained is that users tend to choose graphical methods than descriptive methods [2]. Anton (2009) develops an expert system application with forward and backward chaining methods that can be used to diagnose diseases of rice plants with ease of access for users [3]. Shafinah (2013) conducted a research on methodological trends used in the development of expert systems to manage plant pests and diseases. The results obtained are rule based expert systems more desirable compared to case-based systems (Case-based Reasoning) [4].

The Certainty Factor (CF) method is also used as an inference engine in an expert disease detection system. With this CF method then the detection of symptoms of plant disease will get certainty that the plant is experiencing the actual disease. In this study using...
backward chaining [5]. The case-based reasoning (CBR) method can be used for reasoning in the rice disease identification expert system resulting in 100% sensitivity and an accuracy rate of 82.69% [6]. CBR method is also used in the identification of corn plants with an accuracy rate of 74.63% [7]. Research on expert system to detect plant pests and diseases Dendrobium orchid using Case Based Reasoning (CBR) method with algorithm Similaritas Probabilistic Symmetric. A weighted value of less than 60 percent of the system will not provide suggestions, but the system will store symptom data in the database for a given solution by an expert [8].

The Dempster-Shafer method is also used for the identification of plant pests and diseases that result in a suitability rate of 86.67% [9]. This method can also be combined with the fuzzy method [10].

This study aims to develop a web-based expert system that can identify cassava disease using the method of Case-Based Reasoning and Dempster Shafer Method, and know the comparison of the performance of both methods from the side of identification accuracy. Some theories used in this study are as follows:

1.1 Expert System

Expert systems are computer-based applications that are used to solve problems as the expert thinks. Experts are people who have special skills that can solve problems that can not be solved by the common people [11]. In the structure of the expert system there are two parts, namely the development environment (development environment) and the consultation environment (consultation environment). Development environments are used by expert system makers to build their components and introduce knowledge into the knowledge base. The consulting environment is used by the user to consult with an expert [12].

1.2 Case-Based Reasoning

Case-Based Reasoning (CBR) is a process of remembering a case in the past, then reusing it and adapting it in a new case [13]. The CBR cycle is shown in Figure 1.

The stages of CBR are as follows:
1. Retrieve
   Recover / recover the most similar / similar cases with the new case. This section refers to terms of identification, initial similarity, search and recovery and execution.
2. Reuse
   Reuse (use) information and knowledge of the case to solve the problem. The process of reuse of case solutions obtained in the new context is focused on two aspects: the difference between the previous and the current case, what part of the case has been obtained that can be transferred into a new case.
3. Revise
   Revise (review / refine) proposed solutions.
4. Retain
   Retain parts of the experience that may be useful for solving future problems.

1.3 Nearest Neighbor Similarity

The Nearest Neighbor algorithm is one technique for finding the closest distance of each case (cases) in the database, and how similar the size (similarity) of each source case in the database with the target case. The similarity function in the case is formulated as follows [14]:

\[
\text{Sim}(T,S_l) = \frac{\sum_{i=1}^{n} (t_i - s_{l,i})^2 \cdot w_i}{\sum_{i=1}^{n} w_i}
\]

(1)

Similarities usually is at a value of 0 to 1. A value of 0 means that the two cases are not similar absolute, contrary to the value of 1 case is similar to the absolute. Similarities usually is at a value of 0 to 1. A value of 0 means that the two cases are not similar absolute, contrary to the value of 1 case is similar to the absolute.

1.4 Dempster Shafer Inference

In general the theory of dempster shafer is written in an interval [15]:

\[
[\text{Bel}, \text{Plausibility}]
\]

(2)

Belief (Bel) is a measure of the strength of evidence (symptom) in support of a subset. Has a range between 0 (hence indicating that there is no evidence) until 1 indicates a certainty.

b. Plausibility

\[
\text{Pl}(s) = 1 - \text{Bel}(\neg s)
\]

(3)

Plausibility is also worth 0 to 1. If we are sure of \(\neg s\), then it can be said that Bel (\(\neg s\)) = 1, and Pl (\(\neg s\)) = 0. Plausibility will reduce the level of confidence in the evidence. In the Dempster-Shafer theory we recognize
the frame of discernment denoted by \( \theta \) (theta) and the mass function denoted by \( m \). This frame is the universe of speech from a set of hypotheses so called environment. While the mass function \( (m) \) in Dempster Shafer's theory is the belief level of an evidence measure so it is denoted by \( (m) \). To overcome a number of evidences on Shafer's Dempster theory using a rule better known as Dempster's Rule of Combination

\[
m(\omega) = \frac{\sum_{\omega' \in \Theta} m(\omega') \cdot m(\omega'' \cap \omega')}{1 - \sum_{\omega'' \in \Theta} \sum_{\omega' \in \Theta} m(\omega'' \cap \omega')}
\]

(4)

1.5 Cassava Diseases

The classification of cassava disease in this study consists of 4 diseases and 2 pests where several diseases have the same symptoms [16].

2 Method

2.1 Context Diagram

The system is designed to identify pests and diseases of cassava plants based on data entered by the user. System users are divided into two groups: admin as user manager and expert (expert) as shown in figure 2, while user is a regular user (farmer). The user enters the symptom data encountered by the user. The system will process using case-based reasoning method and Dempster Shafer. The results of identification are then submitted to the user.

Table 1. Case Representation

| ID  | Code | Disease                  |
|-----|------|--------------------------|
| 01  | G01  | Akar Batang Rusak        |
| 02  | G02  | Umbi Rusak               |
| 03  | G03  | Daun menjadi kering      |
| 04  | G04  | Permukaan bawah daun menghimpit cairan |
| 05  | G05  | Daun Berwarna Keabadian  |
| 06  | G06  | Bercak Daun Tidak Teratur |
| 07  | G07  | Daun berubah warna       |
| 08  | G08  | Akar Batang Membusuk     |
| 09  | G09  | Umbi Membusuk            |
| 10  | G10  | Bercak Cekik             |
| 11  | G11  | Lubang Bulat Kecil       |
| 12  | G12  | Jarung daun mati         |
| 13  | G13  | Daun Kering              |
| 14  | G14  | Daun Engat               |
| 15  | G15  | Daun dikeling daerah hijauius |
| 16  | G16  | Penyakit batang tanduk bening pertah |
| 17  | G17  | Mutu ungung              |

Table 2. Weight of Symptoms

| Kode | Nama Gejala | Bobot CBR | Bobot DS |
|------|-------------|-----------|----------|
| G01  | Akar Batang Rusak | 5 | 4.9 |
| G02  | Umbi Rusak | 5 | 4.9 |
| G03  | Daun menjadi kering | 3 | 4.9 |
| G04  | Permukaan bawah daun menghimpit cairan | 3 | 4.9 |
| G05  | Daun Berwarna Keabadian | 3 | 4.9 |
| G06  | Bercak Daun Tidak Teratur | 1 | 4.9 |
| G07  | Daun berubah warna | 5 | 4.9 |
| G08  | Akar Batang Membusuk | 5 | 4.9 |
| G09  | Umbi Membusuk | 5 | 4.9 |
| G10  | Bercak Cekik | 5 | 4.9 |
| G11  | Lubang Bulat Kecil | 1 | 4.9 |
| G12  | Jarung daun mati | 3 | 4.9 |
| G13  | Daun Kering | 3 | 4.9 |
| G14  | Daun Engat | 3 | 4.9 |
| G15  | Daun dikeling daerah hijauius | 1 | 4.9 |
| G16  | Penyakit batang tanduk bening pertah | 3 | 4.9 |
| G17  | Mutu ungung | 3 | 4.9 |

2.2 Case-Based Reasoning Inference

2.2.1 Retrieval Process

The retrieval process used in this study is to compare the symptoms of new cases with symptoms of cases that already exist in the base case, then the results of the comparison will be calculated the level of similarity. In this study, the calculation of similarity score using nearest neighbor. The symptom comparison uses the weight value set by the expert as shown in Table 2.

Sample Case:
The calculation of similarity value of new case (T) with old case (S) is presented in Table 3.

Calculation of similarity use equation (1)

\[
Sim(T, S) = \frac{\sum_{i=1}^{n} w_i \cdot x_i \cdot y_i}{\sum_{i=1}^{n} w_i}
\]

From the above similarity calculation results obtained value of 0.71 or 71%.

2.2.2 Reuse Process

In the reuse process the given identification solution is the solution that has the highest similarity value. For the above case, the highest similarity result is in S1 that is uret pest with similarity value equal to 0,71 or 71%, so the recommended solution is solution for pest uret.
2.2.3 Revise Process

Revisions are part of the system adaptation to cases that have not been identified. Revised case is done by the expert. Experts will revise case / weight based on the symptoms that exist in the base case of the system.

2.2.4 Retain Process

The retain process is the stage that plays a role to manage knowledge management on the system. This stage This stage stores the case into a knowledge base that will be used to solve new cases.

2.3 Dempster Shafer Inference

Calculations with the same case as above use the Dempster-Shafer method. Then each of these symptoms will be tested based on the disease. For the above case tested with rule for Uret pest.

G1 (bel)=0.9; G1(θ)=1-0.9 =0.1
G2 (bel)=0.9; G2(θ)=1-0.9 =0.1
G3 (bel)=0.9; G3(θ)=1-0.9 =0.1
G4 (bel)=0.9; G4(θ)=1-0.9 =0.1

\[
m_3(\theta) = \frac{0.9 \times 0.9 \times 0.9 \times 0.9}{1-(0.1 \times 0.1 \times 0.1 \times 0.1)} = \frac{0.651}{0.9999} = 0.65
\]

3 Result and Discussion

Testing is done by the user directly with the system that has been designed. The test is carried out in two stages, the first stage of testing in accordance with the rule, while the second stage of testing performed using 34 test data. Table 4 shows the result on the first stage that the system using Case-Based Reasoning method was able to identify 100% true cassava plant disease. Whereas with Dempster Shafer method the system only able to identify equal to 75.66%. Recapitulation of the second stage test results is shown in Table 5.

Table 4. Test Result According to Rule

| Case | The Type of Disease   | Value  | CBR | DS |
|------|----------------------|--------|-----|----|
| 1    | Uret                 |        | 100 | 81 |
| 2    | Tungau Merah         |        | 100 | 81 |
| 3    | Hawar Daun Bakteri   |        | 100 | 73 |
| 4    | Layu Bakteri         |        | 100 | 73 |
| 5    | Jamur Akar Putih     |        | 100 | 73 |
| 6    | Bercak Daun Coklat   |        | 100 | 73 |
| **Average** |                        |        | **100%** | **75.66%** |

Test results need to be evaluated to determine whether the system built feasible to be applied in identifying the disease of rice plants. Evaluation is done by calculating the sensitivity and accuracy using equations (5) and (6) [17].

\[
\text{Sensitivitas} = \frac{TP}{TP + FN}
\]

\[
\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN}
\]

The steps taken in testing the system by creating a confusion matrix based on each value of the test results system. Table 6 shows the confusion matrix from the results of the test Table 5.

| Method              | Test Data | TP | TN | FP | FN |
|---------------------|-----------|----|----|----|----|
| Case-Based Reasoning| 34        | 23 | 0  | 11 | 0  |
| Dempster Shafer     | 34        | 11 | 0  | 23 | 0  |

The calculation result of sensitivity value for Case-Based Reasoning and Dempster Shafer is 100% with the system accuracy value for Case-Based Reasoning 67.65% and 32.35% for Dempster Shafer. This shows that the expert system of cassava plant disease identification using Case-Based Reasoning inference can identify cassava plant disease well compared with Dempster Shafer.

4 Conclusion and Suggestion

This research produces expert system of cassava plant disease identification with Case-Based reasoning and Dempster-Shafer inference method. The calculation result of sensitivity value for Case-Based Reasoning and Dempster Shafer is 100% with the system accuracy...
value for Case-Based Reasoning value of 67.65% and 32.35% for Dempster Shafer. This shows that the expert system of cassava disease identification using Case-Based Reasoning inference can identify cassava plant disease well compared to Dempster Shafer. To get more accurate identification result, it is suggested that the weighting is done by some experts and taken on average to get better weighting.

Acknowledgement

The research was funded by the Directorate of Research and Community Service of the Ministry of Research, Technology and Higher Education (Kemristekdikti) for the funding of the National Strategic Research Institute (PSNI) research grant for 2018 under contract number: 324/27.O10.5/PN/II/2018

References

1. Tim Penyusun., Laporan akuntabilitas Kinerja Instansi Pemerintahan (LAKIP) 2015, (Dinas Pertanian Tanaman Pangan Sumatera Barat, Padang, 2015)
2. S.S. Abu-Naser, K.A. Kashkash and M. Fayyad., Journal of Artificial Intelligence, 1, 2 (2008)
3. A. Setiawan, Jurnal Tekkomnika, 7, 3 (2009)
4. Shafinah, K., Sahari, N., Sulaiman, R., Yusoff, M. S. M., & Ikram, M. M., Journal of Theoretical & Applied Information Technology, 58, 1 (2013)
5. M. Minarni, Anisah Fadhillah, Journal of Dynamics, 2, 1 (2017)
6. M. Minarni, I. Warman, Seminar Nasional Aplikasi Teknologi Informasi (SNATI) (2017)
7. M. Minarni, I. Warman, Y. Yusuf, Jurnal Teknoif, 6, 1 (2018)
8. P.A. Aconcagua, S. Wibisono, SINTAK, 1 (2017)
9. Y. Nurcahyo, N. Hidayat, R.S. Perdana, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2, 3 (2018)
10. M. Muliadi, I. Budiman, M.A. Pratama, A. Sofyan, KLIK-Kumpulan Jurnal Ilmu Komputer, 4, 2, (2017)
11. Kusrini, Aplikasi Sistem Pakar, (Andi Offset, Yogyakarta, 2008)
12. Sutojo, T, Edy Mulyanto, Vincent Suhartono. Kecerdasan Buatan, (Andi Offset, Yogyakarta, 2011)
13. Pal, S.K and Shiu, Simon C.K., Foundation of Soft Case-Based Reasoning, (John Wiley &Sons Inc Publication, USA, 2004)
14. Kusrini, Luthfi, E.T., Algoritma Data Mining, (Andi Offset, Yogyakarta, 2009)
15. Kusumadewi, S., Artificial Intelligence (Teknik dan Aplikasinya), (Graha Ilmu, Yogyakarta, 2003)
16. Semangun, H., Penyakit-penyakit Tanaman Pangan di Indonesia (Gadjah Mada University Press, Yogyakarta, 2008)
17. Witten, I.H., Frank,E., Data Mining: Practical Machine Learning Tools and Techniques, 2 (Morgan Kaufmann, San Francisco, 2005)