International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 2 page 518 - 524

Study of C45 Algorithm In Predicting New Employee Acception

Khoirunsyah Dalimunthe[1], Zakarias Situmorang[2]

Email: choyrun@gmail.com
[1]*Universitas Potensi Utama*
[2]*Universitas Katolik Santo Thomas*

ABSTRACT
Good employee performance is one of the criteria in the company where one has good behavior and can complete the work given to him. But there are some difficulties in knowing the quality of people who have good potential as employees in a company. Therefore, a method or method is needed to identify prospective employees of a company. The C4.5 algorithm can be used to predict and classify prospective employees who have the potential to enter the company by making a decision tree based on existing data and predicting new prospective employees who want to enter the company.

Keywords : **C4.5 Algorithm, Candidates for Employees, Predictions, Decision Trees.**

INTRODUCTION

Good and quality employees of course have a positive impact on the companies where they work. Therefore the company must select the people who want to enter and work in the company well. Because employees are the most important resource in a company and good employees and meet company standards can only be obtained through an effective recruitment process (Handy Wicaksono et al., 2008). Acceptance of new prospective employees is a stage where a company recruits people who apply to the company and determines whether the person meets the criteria and needs of the work unit in the company. According to Tjahyono, the reason for the recruitment of new employee candidates is the development of the company's business unit which resulted in the need for additional new employees and the company's need to fill the vacant positions left by their old employees (Li et al., 2016).

In addition, Slamet stated that this recruitment process is useful for obtaining information about the skills, personality, and other abilities possessed by workers who apply to the company. This information is considered very necessary to determine whether the employee is qualified and worthy to enter the company. A qualified workforce will certainly help to improve company performance (Ikhsan et al., 2019). Based on the results of an interview with one of the workers at the Panca Budi Development University, Mr. Marihot, it can be concluded that this company still needs additional assistance in the process of recruiting prospective employees. According to him, there are still some lazy and incompetent office employees. This means that the employee recruitment process at this company is not going well. In addition, based on the results of an interview with Cendrawati as the Human Resource Development (HRD) division, there are several factors that determine whether a person can be accepted into the

International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 2 page 518 - 524

company(Goyal & Kaushal, 2016). These factors are: age, last education, work experience, gender, behavior at interview, request for starting salary and current illness (Lubis, 2018).

The C4.5 algorithm can be used to research various things, including predicting the win rate in soccer matches, looking for rain prediction patterns, to determining the best teacher. Previously, there have also been similar studies using this algorithm, but the attributes used to classify the decision tree are different. Another difference is that the research is conducted on prospective civil servants. The research was conducted by Kumara and Supriyanto with the title "Classification of Data Mining for 2014 Civil Service Candidate Selection Admissions Using the Decision Tree C4.5 Algorithm". The level of accuracy obtained using the C4.5 algorithm is quite high (Li et al., 2016),

The C4.5 algorithm is used in this study to predict the process of accepting new employees at the Panca Budi Development University. According to research by HSSINA, et al, the C4.5 algorithm is the strongest algorithm for the decision tree making process when compared to other algorithms such as ID3, C5.0, and CART. Based on this research, the C4.5 algorithm was decided to be used in this study because this algorithm is stronger than other algorithms. Data on employees who work as field employees (maintenance, service) will be used as training data for making a decision tree that will be used to predict employee acceptance and test the success rate of these predictions using data testing on employees who have the same job position.

LITERATURE REVIEW

Some of the literature that underlies this research will be discussed in this chapter, including decision support systems, decision trees, C4.5 algorithms, and cross-validation (Aryza et al., 2018).

Decision Support System.

Decision support system is a system that can help humans to make decisions objectively. The concept of a system like this was first coined in the 1970s by Michael S. Scott Morton, Michael first called a system likeThis is called the Management Decision System. The purpose and objective of the decision-making system is to support decision-makers to choose alternative decisions using decision-making models and to solve problems that are structured, semi-structured, or unstructured (Machfoedz, 2015).

According to Mengkepe, the basic framework for managerial decision making in this type of decision is divided into several parts, namely:

1. Structured: Contains problems that often occur, the solution can be standard and standard.
2. Unstructured: Contains complex problems using non-standard problem solving, the solution of which involves human intuition as the basis for decision making.
3. Semi Structured: A combination of structured and unstructured decisions, the solution is a combination of standard solution procedures with individual human capabilities.

International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 2 page 518 - 524

Decision Tree.

Humans are always faced with various kinds of problems from various fields of life. This problem also has varying degrees of difficulty (Goyal & Kaushal, 2016). To deal with this problem humans began to develop a system to help them solve these problems, one such system is a decision tree. Decision tree is a classification and prediction method that has proven to be powerful and very well known. This method serves to convert facts into decision trees that represent rules that can be easily understood in natural language. The process of this decision tree starts from the root node to the leaf node which is carried out recursively where each branch represents a condition and each end of the tree will represent a decision (Muhathir & Al-Khowarizmi, 2020).

The decision tree architecture is made in such a way as to resemble the original tree, where there are several parts, namely:

- *Root Node*: This node is located at the very top of the decision tree.
- *Internal Nodes*: This node is a branch which requires one input and produces a maximum of two outputs.
- *Leaf Nodes*: This node is a node located at the end of the tree. This node only has one input and no output.

C4.5 . Algorithm

The C4.5 algorithm is an algorithm used to perform the data classification process using a decision tree technique. The C4.5 algorithm is an extension of the ID3 algorithm and uses a similar decision tree principle. This algorithm is very well known and preferred because it has many advantages. These advantages, for example, can process numeric and discrete data, can handle missing attribute values, produce rules that are easy to interpret and their performance is one of the fastest compared to other algorithms.

The basic idea of this algorithm is making a decision tree based on the selection of the attribute that has the highest priority or can be called the highest gain value based on the entropy value of the attribute as the axis of the classification attribute. Then recursively the tree branches are expanded so that the whole tree is formed. According to the IGI Global (International Publisher of Progressive Academic) dictionary, entropy is the amount of data that is irrelevant to information from a data set. Gain is information obtained from changes in entropy in a data set, either through observation or it can also be concluded by participating in a data set.

Based on what was written by Jefri, there are four steps in the process of making a decision tree in the C4.5 algorithm, namely:

1. Selecting attribute as root
3. Create a branch for each value
4. Split each case in a branch
5. Repeats the process in each branch so that all cases in the branch have the same class.

According to Jiandi the data owned must be compiled into a table based on cases and the number of respondents before calculating the entropy and gain values. used to determine how informative the attribute is. Here's the statement:

International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 2 page 518 - 524

```
FormTree(T)
        (1) ComputeClassFrequency(T);
        (2) if OneClass or FewCases
                Return a leaf;
            Create a decision node N;
        (3) ForEach Attribute A
                ComputeGain(A);
        (4)  N.test = AttributeWithBestGain;
        (5) if N.test is Continuous
                Find Threshold;
        (6) ForEach T1 in the Splitting of T
        (7)     if T1 is Empty
                        Child of N is a leaf
                Else
                        Child of N = FormTree(T1);
        (8) ComputeErrors of N;
            Return N;
```

Figure 1. Pseudocode Algorithm C4.5

Figure 1 is the pseudocode of the C4.5 algorithm which functions for the formation of a decision tree. The calculation starts from counting the number of attributes and determining which attribute will be used as the root of the decision tree. Furthermore, entropy and gain calculations will be carried out to determine the leaf of the decision tree. After all calculations are completed, a decision tree can be formed based on the previously calculated gain value. The attribute with the highest gain value will be in a higher priority and have a higher position in the decision tree.

Cross Validation

Cross validation is a statistical method used to evaluate and compare a data set by dividing the data into two parts, namely training data and testing data. One type of cross validation is ten-fold cross validation. This validation is done by dividing a data set into ten segments– 10the same size by randomizing the data. Then1will be used first for the training process and validated using the rest of the data other than1. After that it will be used for training, while the rest of the data other than2used for validation, and so on. By doing validation like this, the accuracy that will be obtained will be higher.

METHODS

The implementation of the C4.5 algorithm for predicting the acceptance of new recruits uses the following research steps:

International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 2 page 518 - 524

1. Literature Study
   The research begins by studying the information and algorithms related to this research by reading e-books, e-journals, and several other learning references. At this stage the concepts needed in the research will be finalized, such as the definition of the C4.5 algorithm and its application.
2. Sample Collection
   Data At this stage the process of collecting data samples will be carried out by asking for employee data directly from the company, namely PT WISE. This employee data includes various kinds of information ranging from name, age, salary, address, and other attributes which will then be processed using the C4.5 algorithm. The amount of data that will be used is 84 field employee data.
3. Data Sample Analysis
   After the employee data is obtained, the data attribute is sorted and calculated according to the predetermined parameters to calculate the entropy and gain values to get an overview of a data set.
4. Application Design and Development
   At this stage, several things will be determined, namely what procedures and processes can be carried out by the application, the process flow, and the basic appearance of the application. The application design will be represented in the form of a diagram that will describe the process flow of the application. After that, the application starts to be built using the right programming language.
5. Test Application
   At this stage, trials will be carried out on applications that have been made on the data that has been collected previously. Observations on whether the C4.5 algorithm can be implemented properly on the system and can have a high level of accuracy in the prediction process for new employees are also carried out at this stage.
6. Analysis of Application Results
   After the test on the application has been successfully carried out, the next step will be to measure the level of prediction accuracy using ten-fold cross validation. Cross validation is used to make the prediction measurement accuracy more precise.
7. Report Writing
   At this stage, the process of recording every activity carried out during this research process takes place and distributes the information in the form of a report as a form of documentation.

RESULTS AND DISCUSSION

This section describes the results of measuring the accuracy of applications that have been built in the study. The trial was carried out using 84 sample data and using the ten fold cross validation measurement method. The data will then be divided into 10 groups of training data and testing data. After that, a trial to measure the accuracy of the application is carried out. The analysis that will be carried out on the application is the calculation of the level of accuracy using the 10-fold cross validation method, both per cluster accuracy and overall. The trial will be carried out on 84 samples of data in the

International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 2 page 518 - 524

database automatically through the system.

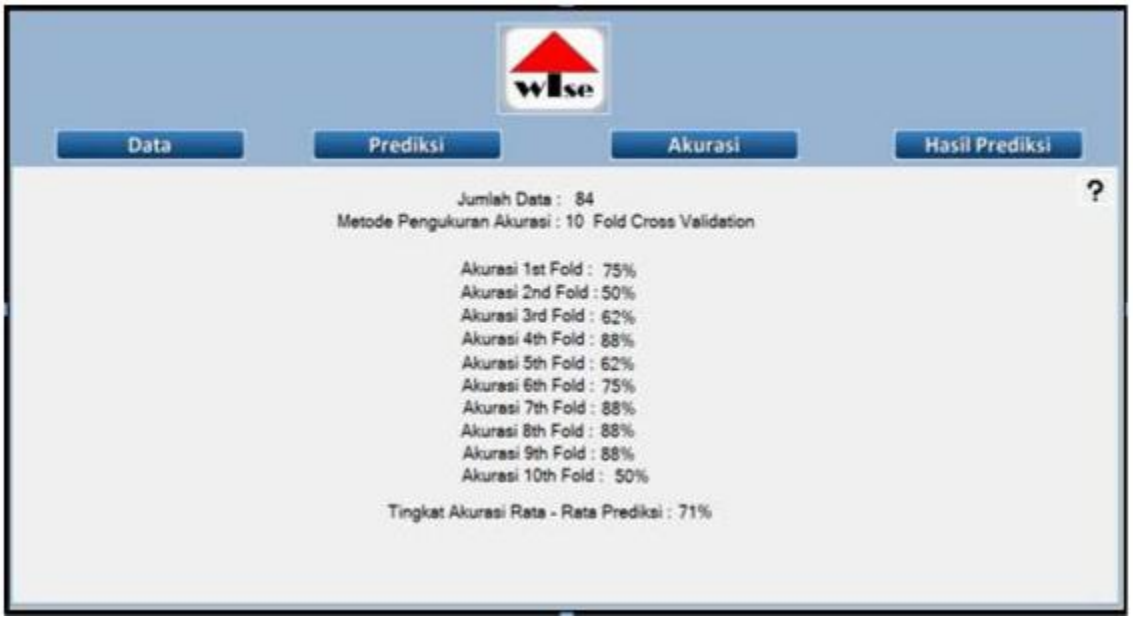

Figure 2. Accuracy Measurement Results

Figure 2 describes the results of calculations from cross validation carried out by the application. Can be seen the results of the calculation of accuracy for each cluster and as a whole. The overall accuracy measurement result is 71%. The prediction results do not reach 100% because there are several cases where the data sample is still not large enough so that the prediction results are still ambiguous. These additional samples can update the tree thereby reducing the prediction error rate.

CONCLUSION

The implementation of the C4.5 algorithm to predict prospective new employees at PT WISE has been successfully carried out. The result of the prediction of the success rate of prospective new employees at PT WISE as a whole which has been measured using the ten-fold cross validation method is 71%. For further research, it is advisable to conduct trials using more data samples so that the level of application accuracy can be increased.

REFERENCES

Aryza, S., Irwanto, M., Khairunizam, W., Lubis, Z., Putri, M., Ramadhan, A., Hulu, F. N., Wibowo, P., Novalianda, S., & Rahim, R. (2018). An effect sensitivity harmonics of rotor induction motors based on fuzzy logic. *International Journal of Engineering and Technology(UAE)*, *7*(2.13 Special Issue 13), 418–420. https://doi.org/10.14419/ijet.v7i2.13.16936

International Journal of Economic, Technology and Social Sciences
e-ISSN 2775-2976
International Journal of Economic, Technology and Social Sciences
url: https://jurnal.ceredindonesia.or.id/index.php/injects
Volume 2 Number 2 page 518 - 524

Goyal, R. K., & Kaushal, S. (2016). Effect of utility based functions on fuzzy-AHP based network selection in heterogenous wireless networks. *2015 2nd International Conference on Recent Advances in Engineering and Computational Sciences, RAECS 2015*, *December*, 0–4. https://doi.org/10.1109/RAECS.2015.7453366

Handy Wicaksono, Resmana Lim, & William Sutanto. (2008). Perancangan SCADA Software dengan Wonderware InTouch Recipe Manager dan SQL Access Manager pada Simulator Proses Pencampuran Bahan. *Jurnal Teknik Elektro*, *8*(1), 38–45. https://doi.org/10.9744/jte.8.1.38-45

Ikhsan, M. G., Saputro, M. Y. A., Arji, D. A., Harwahyu, R., & Sari, R. F. (2019). Mobile LoRa gateway for smart livestock monitoring system. *Proceedings - 2018 IEEE International Conference on Internet of Things and Intelligence System, IOTAIS 2018*, 46–51. https://doi.org/10.1109/IOTAIS.2018.8600842

Li, Z., Shang, C., & Shen, Q. (2016). Fuzzy-clustering embedded regression for predicting student academic performance. *2016 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2016*, 344–351. https://doi.org/10.1109/FUZZ-IEEE.2016.7737707

Lubis, A. H. (2018). *ICT Usage Amongst Lecturers and Its Impact Towards Learning Process Quality*. *34*(1), 284–299.

Machfoedz, M. M. (2015). Stabilizing and Decentralizing the Growth through Agro-industrial Development. *Agriculture and Agricultural Science Procedia*, *3*, 20–25. https://doi.org/10.1016/j.aaspro.2015.01.006

Muhathir, & Al-Khowarizmi. (2020). Measuring the Accuracy of SVM with Varying Kernel Function for Classification of Indonesian Wayang on Images. *2020 International Conference on Decision Aid Sciences and Application (DASA)*, 1190–1196. https://doi.org/10.1109/DASA51403.2020.9317197