

DATA MINING-BASED PLACEMENT OF COMPETENT EMPLOYEES USING DECISION TREE WITH THE C4.5 ALGORITHM

¹EDI KARNOTO, ²SURYARINI WIDODO

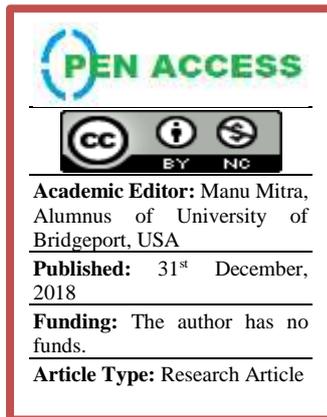
^{1,2}Master of Technology and Engineering Study Program, Gunadarma University

^{1,2}Jl. Margonda Raya No. 100, Depok West Java, INDONESIA - 16424

Email: ¹edi.karnoto@gmail.com, ²srini@staff.gunadarma.ac.id

ABSTRACT

The Personnel Management Organizing Body is an agency that deals with government affairs that its duties in the field of civil service management are in accordance with the provisions of the legislation. In carrying out governmental duties in the field of personnel management, the Personnel Management Organizing Body requires qualified Apparatus, Employees and their placement must be in accordance with their competence. Until now there are still problems with the placement of employees in positions that are not in accordance with their competence. Data mining; in this study was implemented to measure the level of suitability of employees for positions occupied by involving large amounts of data, the technique used for classification is the decision tree with the C4.5 algorithm. The C4.5 algorithm is used as a tool to determine the extent of suitability between employees and positions they occupy. The parameters that are used consist of five elements namely education, training, job experience, expertise, and skills. With the object of the study using 540 profile data of



employees who hold administrator positions (equivalent to echelon III, 130 people) and supervisors (equivalent echelon IV, 410 people). The accuracy, of suitability between employees and positions they occupy is based on testing of the results of the classification of the decision tree method and C4.5 algorithm, which is conducted using 90% training data and 10% testing data, is 95 %.

Keywords : personnel management organizing agency; employee placement; data mining; decision tree; c4.5 algorithm; civil service

1. INTRODUCTION

Human Resources (HR) Apparatus is a valuable asset and has an important role in realizing the vision and mission and goals of a government organization. The goal of government organizations can be achieved optimally if it is supported by qualified HR personnel. The quality of HR Apparatus in a Government Agency depends on the personnel management process carried out in the agency. As for there are still some problems in personnel management, among them are still low employee competency and not in accordance with the needs or placement in the position so that it is less optimal in carrying out job duties so that the planned target is not reached optimally.

Ali Sutrisno in 2015 conducted a study that made recommendations for employee promotion using the decision tree method with a C4.5 algorithm. This method provides promotion recommendations with better accuracy than other algorithms. In addition, the Decision Tree classification with C4.5 algorithm has a high degree of accuracy, both of which have been formulated into applications and Weka, each level of accuracy is 100% and Weka is 98.67% [1]. By using the same method, in 2017 Arief Hidayat conducted research by applying data mining to predict the level of employee competence at PT, Hankook Tire Indonesia. Testing was carried out on 164 employees with a fairly high accuracy rate of 90.7%. This research category is classified as very good [2].

So far, the Personnel Management Organizing Body has conducted a competency assessment to measure the suitability between the profile of employees and the positions they occupy by carrying out an assessment of each employee who occupies a position, where the assessment process requires significant costs and requires facilities and infrastructure. Based on the background described above, the researcher tries to apply the decision tree method with a C4.5 algorithm as a tool to measure the suitability of employee placement. Measurement of the suitability of placement of employee is carried out by taking into account the five elements namely education, training, experience, expertise, and skills. The study was conducted on 540 profile data of employees who hold administrator positions (equivalent to echelon III, 130 people) and supervisors (equivalent to echelon IV, 410 people).

2. LITERATURE STUDY

2.1 Data mining

Data mining is a process that employs one or more machine learning techniques to analyze and extract knowledge automatically [5]. Data mining is a term that is often said to be a way to describe and find discoveries in the form of knowledge in a database. Data mining is the process of selecting or "mining" knowledge from a large number of data sets [5]. In general, there are two types of methods in Data mining, namely:

1. The Predictive method processes to find patterns of data that use several variables to predict other variables that are unknown in type or value. The techniques included in predictive mining include Classification, Regression, and Deviation [1].
2. The Descriptive method processes to find an important characteristic of data in a database. Data mining techniques included in descriptive mining are Clustering, Association, and Sequential Mining [1].

The stages in data mining can be seen in Figure 1 below:

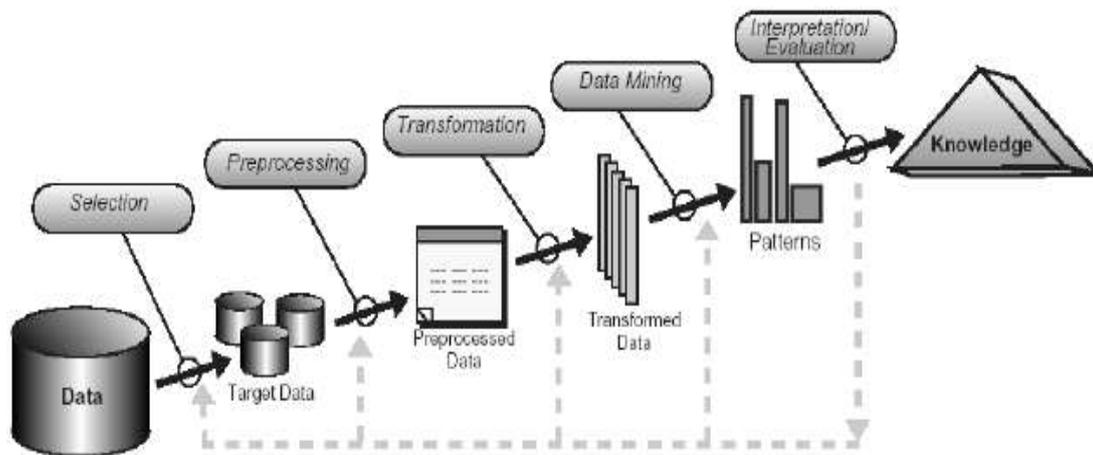


Figure. 1 Steps in data mining [6]

The stages in data mining are as follows:

1. Data cleaning is a process of eliminating inconsistent noise and data or irrelevant data.
2. Data integration is a combination of data from various databases into a new database.
3. Data selection, only the data that is suitable for analysis will be taken from the database.
4. Data transformation, changing or merging data into a format suitable for processing in data mining.
5. Mining process, the main process to find valuable and hidden knowledge from data.
6. Evaluate patterns, to identify interesting patterns into knowledge-based.
7. Presentation of knowledge is a visualization and presentation of knowledge about the methods used to obtain knowledge obtained by users.

The data used in this study consists of primary data and secondary data. Primary data is obtained by interviewing the Administrator Officer, namely the Head of Personnel Development Section, including the first, the mechanism of placement/appointment of employees in an office within the Civil Service Management Agency. Second, the difference between appointment/placement at each level of office and the third clarification related to the method of appointment/placement of employees that are currently running. Secondary data is obtained from archives, literature and other sources related to the problem being studied. As a research material for employee profile data and job title data obtained from the Managing Agency for Manpower Management electronically and manually.

2.2 Classification

Data classification is a process that finds the same properties in a set of objects in a database and classifies them into different classes according to the specified classification model [1]. The classification technique is divided into several techniques which are described in Figure 2.

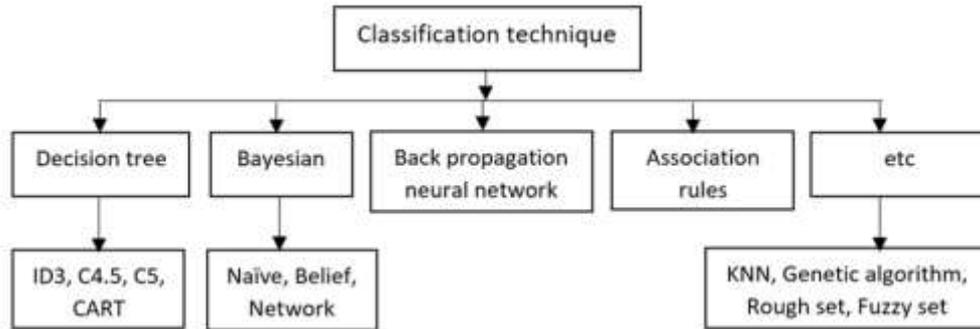


Figure. 2 Classification technique [1]

2.3 Decision tree

The decision tree is a flow-chart like a tree structure, where each internal node shows a test on an attribute, each branch shows the results of the test, and leaf node shows classes or class distribution [1].

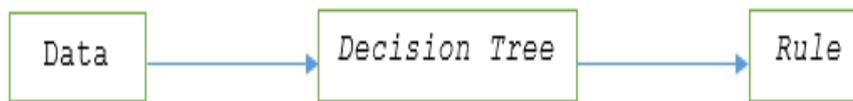


Figure. 3 Basic concepts of decision tree [1]

The sample data is made into the decision tree using certain algorithms which will then be made Rule (provision) to predict new data [1].

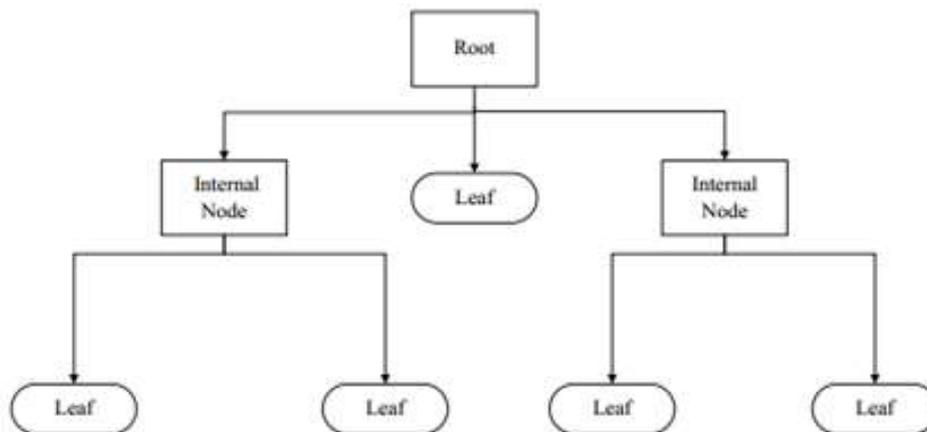


Figure. 4 Example decision tree [1]

2.4 C4.5 algorithm

The C4.5 algorithm is a method used to make a decision tree that has input in the form of training samples and samples. Training samples are sample data that will be used to build a tree that has been tested for truth. Samples are data fields that will be used as parameters in classifying data [3]. The work steps of the C4.5 algorithm can be described as follows:

1. Calculate Entropy from every a tribute with use formula [4] :

$$\text{Entropy}(S) = -p+\log2p+ -p.\log2p..(1)$$

With the following information:

- S = sample data space used for training.
- P + = amount of positive support bersolusi on sampled data u ntuk certain criteria.
- P - = number of negative polls that do not support sample data for certain criteria.

2. Information Gain of each attribute using the formula [4] :

$$\text{Gain (S,A)} = \text{Entropy (S)} - \sum_{V \in \text{val}} \frac{|S_V|}{|S|} \text{Entropy}(S_V).. (2)$$

With information as the following:

S = space sample data used for training.

A = attribute

V = the value of the possible for attribute A.

Value (A) = set values which may for attribute A.

|S_v| = amount sample for the value of V.

|S| = amount all over sample data.

Entropy (S_v) = entropy for samples have the value of V.

3. Gain Ratio of each attribute uses the formula [4] :

$$\text{Gain ratio (S,A)} = \frac{\text{Gain (S,A)}}{\text{Split Info (S,A)}} \dots(3)$$

The purpose of calculating Information Gain and Split Info is to get the Gain Ratio value. The attribute in the Decision Tree can be determined based on the value of Gain Ratio [4].

2.5 WEKA

Weka (Waikato Environment for Knowledge Analysis) is a machine learning software to solve popular real-world data mining problems written in Java and runs on almost all platforms, developed at the University of Waikato, New Zealand [8]. Weka supports several data mining standard tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection [7].

2.6 Object of research

The object of this research is employee profile data and job requirements data which consists of 5 elements including education, training, experience, skills and skills that hold an administrator position (echelon III equivalent) and supervisor (echelon IV equivalent) with a dataset administrator officials as many as 130 people and datasets of supervisory officers 410 people.

3. METHODOLOGY

Six phases of data mining development cycle are shown in Figure 5.

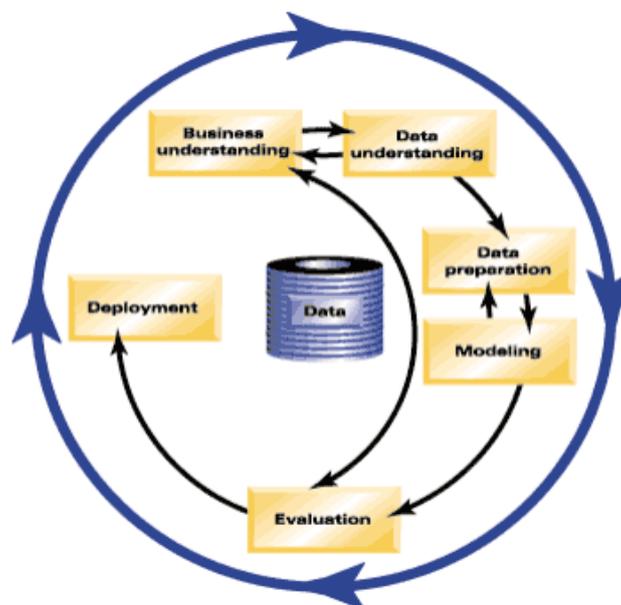


Figure. 5 Stages of the research process in CRISP-DM [5]

The stages of the research process in CRISP-DM that will be carried out are explained as follows:

3.1 Understanding business

Every employee on agency government is appointed in position Employee Country Civil. The existing jobs on agency government have assignments that describe core tasks (core business) from the agency so that in filling it the key focus must be on education, training, experience, expertise and skills that must be fulfilled in order for the task position run optimally.

3.2 Understanding data

Primary data is obtained directly from location research where the author hold data collection through the interview with the Head Part Development Employment Bureau work units, by means of submitting some questions related to staffing. Secondary data is obtained from documents and files existing in the company in relation to this research, or the data that was obtained from various literature and other sources related to issues being researched.

3.3 Data processing

In the data processing data is sorted as shown in Table 1.

Table. 1 Attribute selection

Attribute	Usage Details	
Employee ID number	X	No
Name	X	No
Position	√	Model Value
Work unit	X	No
Education	√	Model Value
Education and Training	√	Model Value
Experience	√	Model Value
Expertise	√	Model Value
Skills	√	Model Value

3.4 Modelling

The proposed research model can be illustrated in Figure 6 below:

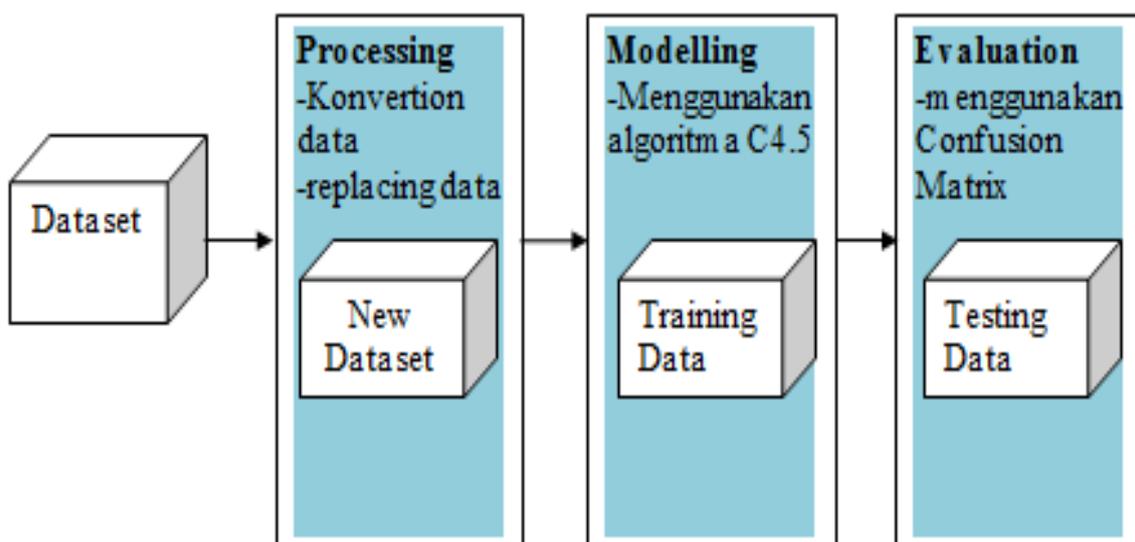


Figure. 6 The proposed research model [6].

At this stage, the modeling process is carried out on datasets, namely processing and modeling with C4.5 algorithm and evaluation. Processing consists of two stages, namely data conversion and data replacing. Data conversion results can be seen in Table 2.

Table. 2 Data conversion results

Attribute	Data Type	Range of Data /Description
Education	Text	Corresponding / Not yet Corresponding
Education and Training	Text	Corresponding / Not yet Corresponding
Experience Work	Text	Corresponding / Not yet Corresponding
Knowledge Work	Text	Corresponding / Not yet Corresponding
Skills Work	Text	Corresponding / Not yet Corresponding
Results	Text	Corresponding / Not yet Corresponding

Replacing data is done by validating each attribute such as deleting data from administration officials and supervisors who are employed in other agencies.

3.5 Validation and evaluation

In this step, the process is validated as well as accuracy of the results is measured. As a measurement tool, this research uses WEKA software.

3.6 Deployment

The results of this study are in the form of analysis leading to the Decision Support System (DSS), which is expected to be used by the Personnel Management Organizing Body in filling the position of the Civil Servants and can also be used as reference material for further research.

4. RESEARCH RESULTS AND DISCUSSION

The results of this study are obtained by applying the standard data mining process of the CRISP-DM model (Cross Industry Standard Process For Data Mining) along-with the proposed modeling process as follows:

4.1 Data Conversion

To simplify the classification technique, the conversion process is carried out by comparing the profile attributes of employees with the attributes of job requirements to the position gap attribute dataset. The results of data conversion in each attribute are in the form of text with the following results:

4.1.1 Administrator position

Administrator position conversion results in the form of datasets obtained from benchmarking on each employee profile attributes and job requirements which can be seen in Table 3.

Table. 3 Dataset administrator position

Position	Education	Education and Training	Experience	Expertise	Skills	Prediction Results
Head of Regional I Appointment and Pension Division	Corresponding	Corresponding	Corresponding	Corresponding	Not yet Corresponding	Corresponding
Head of Program and Budget Accountability Section	Corresponding	Not yet Corresponding	Not yet Corresponding	Corresponding	Corresponding	Corresponding
Head of Organization and Management Section	Corresponding	Corresponding	Corresponding	Corresponding	Corresponding	Corresponding

4.1.2 Supervisor position

Administrator position conversion results in the form of datasets obtained from benchmarking on each employee profile attributes and job requirements are shown in Table 4.

Table. 4 Dataset supervisor position

Position	Education	Education and Training	Experience	Expertise	Skills	Prediction Results
Head of Section Award	Corresponding	Not yet Corresponding	Not yet Corresponding	Corresponding	Corresponding	Corresponding
Head of Section Verification and Distribution of Personnel Archives	Corresponding	Not yet Corresponding	Not yet Corresponding	Not yet Corresponding	Not yet Corresponding	Not yet Corresponding
Head of Personnel Status Section	Corresponding	Not yet Corresponding	Corresponding	Corresponding	Corresponding	Corresponding

4.2 Replacing Data

Replacing data is carried out on the results of data conversion as shown in Table 5. The results in Table 5 explain that the number of data used is 137 administrator officials and 436 supervisory officials, but later the data is replaced into 130 administrator officials and 420 supervisory officials. Replacing data by validating each attribute that is used, such as deleting data from administration officials and supervisory officers employed at other agencies with each amount of data as many as 7 administrator officials and 16 supervisory officials.

Table. 5 Results replacing data

Position level	Results conversion	Replacing results
Administrator	137	130
Supervisor	436	420

4.3 Modelling

To find out the extent of the validity of the classification results, it is necessary to carry out testing. Testing is performed three times against the results of the classification with a different percentage of training data and testing data, the best results are obtained by testing using 90% training data and testing data 10%.

4.4 Evaluation

The level of accuracy obtained from the classification results is measured using a confusion matrix. The data generated from the confusion matrix explain the number of prediction errors in the classification technique for training data and testing data. Measurement of the level of accuracy with the confusion matrix is done for each test result. The results of the classification of training data and testing data are in the form of a classifier model in the form of a decision tree. Classification results and predictions in each position are as follows:

4.4.1 Classification results of administrator official dataset

The results of testing the administrator's office testing data obtained prediction results as shown in Table 6.

Table. 6 Testing results

Data Training	Data Testing	Akurasi	Precision	Recall	F-Measure Recall
117	13	92,3%	85,2%	92,3%	88,6%

From the test results obtained accuracy of 92.3%, with confusion matrix on 13 employee data there was only 1 prediction error, namely 1 employee who was not in accordance with his position. Decision tree pattern of the suitability of employees to the position is not formed in the classification process.

4.4.2 Classification results of supervisor official dataset

The results of testing the supervisory office testing data obtained the prediction results as shown in Table 7.

Table. 7 Testing results

Data Training	Data Testing	Akurasi	Precision	Recall	F-Measure Recall
369	41	99,7%	97,7%	97,6%	97,6%

From the test results obtained Accuracy of 92.3%, with confusion matrix that 41 data only occurred 1 prediction error on 1 employee who was not in accordance with his position. The decision tree for employee suitability for their position is shown in Figure 7.

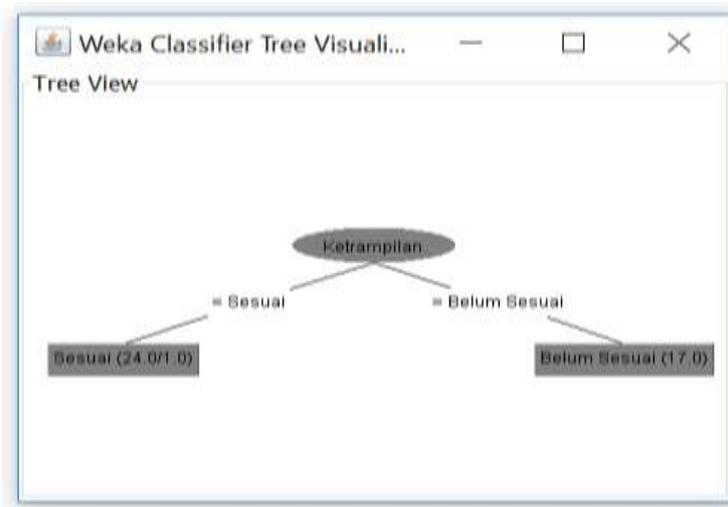


Figure. 7 Tree testing

4.4.3 Predicted results

Based on the results of the testing of the dataset Administrator and Supervisor officer, the prediction results are shown in Figure 8.

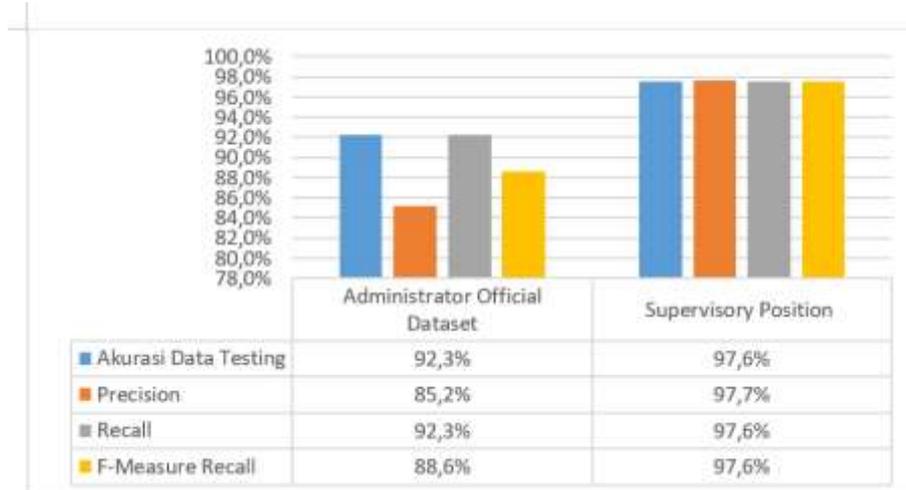


Figure. 8 Predicted results

Classification of accuracy level measurements is divided into several groups as shown below:

- a. 0.90-1.00 = very good classification
- b. 0.80-0.90 = good classification
- c. 0.70-0.80 = sufficient classification
- d. 0.60-0.70 = bad classification
- e. 0.50-0.60 = incorrect classification

Taking into account the value of accuracy obtained in the classification technique, it is in accordance with the classification of the accuracy values generated in the application of the classification technique with the C4.5 algorithm into a very good classification category.

5. CONCLUSION

The prediction results explain that the testing of the results of the classification of the decision tree method and the C4.5 algorithm is carried out using 90% training data and 10% testing data. Accuracy of the level of suitability between employees with positions occupied based on the position gap measurement obtained 95%, so

that it has a very good classification category which is known that if one attribute of the terms of the job are skills and nine pairs of job requirements are skills, expertise and experience, training and skills, experience and training, expertise, education and training, expertise, skills and education, expertise, skills and training, skills, skills and experience, experience, skills and education accordingly, the employee profile is in accordance with the position occupied. Thus, ten pairs of the attributes can be used to predict the suitability of employees to the positions occupied. Educational attributes which are the main attributes after testing with the classification technique with C4.5 algorithm do not become the main element for employees to occupy a position, especially administrator positions and supervisory positions.

ACKNOWLEDGMENT

Alhamdulillah Rabil 'Alamin, the researcher expressed his highest gratitude to Allah subhanahu Wa ta'ala for blessings, love, opportunity, health, and mercy to complete this research. The authors say thanks and appreciation to the Personnel Management Organizing Body as the agency that has allowed to conduct research there. Appreciation of the Head of Personnel Development Section of the Personnel Management Organizing Agency who has the pleasure to provide data and information as the material for this study. The results of this study are far from perfect, but it is hoped that it will be useful not only for researchers but also for readers. For this reason, constructive suggestions and wise criticism are desirable. Finally, may Allah subhanahu Wa ta'ala repay all the goodness and help that has been given with multiple rewards. Hopefully, the results of this study provide benefits for the development of science in general and for the author in particular and can be a recommendation or alternative for the Personnel Management Agency as a reference for the placement of employees in the administrator position and supervisory positions in the future.

REFERENCE

1. Ali Sutrisno, Lely. Hiryanto., Dyah E. Herwindiati, *Employee Position Increase Recommendation System Using Decision Tree Method With C4.5 Algorithm (Case Study of Pt Cipta Saksama Indonesia)*. National Information Technology Seminar (NITS), Tarumanegara University, Jakarta, 2015.
2. Arief Hidayat, A. S. S., *Application of Data Mining for Predicting Employee Competency Levels at Pt. Hankook Tire Indonesia With C4.5 Algorithm Method*. National Seminar on Information and Communication Technology, Atmajaya University, Yogyakarta, 2014.
3. Lutfi, K. E. T., *Data Mining Algorithm*. Yogyakarta: Andi Offset, 2009
4. Mochammad Kautsar Sophan, Rika Yunitarini, Windy Julianto, *C4.5 Algorithm for Employee Performance Assessment*. Trunojoyo Madura University, ISSN: 1978-0087, Scan Vol. IX Number 2 June 2014.
5. Rendragraha Kumara., Catur Supriyanto, *Classification of Data Mining for Receiving Selection of Prospective Employees of 2014 Civil Servants Using the Decision Tree C4.5 Algorithm (Case Study: Acceptance of Selection of 2014 Civil Servants at the Regional Civil Service Agency of Banjarnegara District Using the Decision Tree C4.5 Algorithm)*, University Dian Nuswantoro Repository, Semarang, 2015.
6. Sigit Hadi Waluyo., Prihandoko. *Classification of Use of the Prosperous Rice Program (RASTRA) Based on the Level of Poverty Using the Decision Tree C4.5 Algorithm Based on Particle Swarm Optimization*. Energy Scientific Journal of the Engineering Sciences of the University of Panca Marga, Probolinggo, ISSN: 2088-4591, Vol.7 No.2 November 2017 Edition.
7. Yudi Wibisono, *Weka Tutorial*. Bandung: Computer Science University of Indonesia Education, 2017

APPENDIX

Appendix 1

The data used as research material is employee profile data and job requirements data. Employee profile data consists of employee name, position name, education, training, experience, expertise, and skills obtained from the employment application system which can be explained that education data is obtained from the education column from the education table, training data is obtained from the column_name from the course table, occupational experience data is obtained from the nama_jabatan column in the position table, expertise data is not clearly seen in employee profile data in the application system, so the education department and the type of training that employees have can be converted as expertise areas, and skills data is also clearly not visible. in employee profile

data in the application system, the types of training and job experience that employees have can convert the skills they have.

While the data on job requirements are obtained from the job analysis document at point 15 concerning the job requirements that have been prepared by the staffing division of the personnel management body which can be explained that the education data is obtained from the letter b. education that explains the level of education requirements, the education data is obtained from the letter c. education and training that explains the level of training requirements that must be followed, job experience data obtained from the letter d. work experience that explains the position experience that must be possessed, expertise data obtained from letter e. work knowledge that can be converted as a must-have skill, and skill data is obtained from letter f. Job Skills that explain skills that must be possessed. The points for obtaining data and information on terms of office can be seen as follows.

Appendix 2

Transcripts of interviews or surveys conducted in research. Questions and answers from respondents can write as follows:

a. Q (Researcher):

What is the mechanism for placement/appointment of employees in a position within the Civil Service Management Agency?

A (Resource Person):

In accordance with the State Civil Apparatus Law Number 5 of 2014 and its derivatives Government Regulation Number 11 of 2017 mandates that in appointing employees to a position must be based on a merit system which means that in appointing employees based on qualifications, competencies and performance fairly and fairly without distinguish political background, race, color, religion, origin, gender, marital status, age, or disability condition. In relation to the appointment process for the position level, the position of middle leader and pre-service leadership position is done by means of open selection, namely to open a wide selection of registrations that can be followed by officials who have met the requirements in accordance with Government Regulation Number 11 of 2017, while the administrator rank , supervisors through psychological test selection, computer-based Basic Competency Test and assessment.

b. Q (Researcher):

In terms of position, is there anything that distinguishes appointment/placement in an office?

A (Resource Person):

There are, in accordance with the laws and regulations that certainly exist, in terms of qualifications, each echelon position level is definitely different, for example, first for echelon 3 (administrators) has a minimum qualification of undergraduate or diploma IV education and echelon 4 (supervisor) has a minimum qualification of diploma III education or equivalent, both have job experience in the field of assignment that will be occupied and the third has competence as evidenced by the results of the evaluation by the performance assessment team.

c. Q (Researcher):

When viewed from the development of existing regulations, it may be said at this time that the requirements for the appointment/placement of employees in a position have begun to have a measure such as educational requirements, experience, and competencies that must be possessed. Is it true that it is related to education and experience, education can be seen from the last diploma possessed and experience can be seen from the position of the employee before? Is it true that matters related to competency need to be made an assessment of the employee and of course it costs a lot of money?

A (Resource Person):

Yes, right.

d. Q (Researcher):

Is there any impact on the performance of the organization from the results of the selection for the placement of employees in a position.

A (Resource Person):

Of course, there is one example is that there are still employees stationed in positions that do not describe previous job experience so as to adapt to the new job position for a long time and even today there are still some who do not understand what the work is like so that the work of the unit is lacking maximum.

e. Q (Researcher):

As I have said before and heard the description that you conveyed, can I take conclusions to mean that it is possible to evaluate the placement of employees and if allowed on this issue will be appointed as a research for my thesis during the study permit?

A (Resource Person):

Please, and that's what we need.

AUTHORS PROFILE



Edi Karnoto, S.Kom earned a Bachelor's degree in Computer from Budi Luhur University, Jakarta, Indonesia in 2013. Currently working in the State Personnel Board as a Data and Information Analyst, and is currently pursuing a Masters in Information Systems education at Gunadarma University, Depok.



Dr. Suryarini Widodo earned a Bachelor of Computer Science from the University of Indonesia, Depok, in 1996 and obtained a Masters in Information Systems Management from Gunadarma University, Depok, Indonesia in 1998. He earned a Doctorate in Information Technology from Gunardarma University, Depok, Indonesia in 2009. When this works as the Head of Internet Development Center and as a Lecturer at Gunardama University, Depok, Indonesia. His research interests are Pattern matching in Text with Brute-Force Algorithm and Boyer-Moore Algorithm.