

Decision Support Model for Employee Selection Based on Data Mining and Fuzzy Logic

Clarissa Elfira Amos Pah and Ditdit Nugeraha Utama

Computer Science Department, BINUS Graduate Program - Master of Computer Science,
Bina Nusantara University, Jakarta, Indonesia 11480

Article history

Received: 22-01-2020

Revised: 21-04-2020

Accepted: 03-06-2020

Corresponding Author:
Clarissa Elfira Amos Pah
Computer Science Department,
BINUS Graduate Program -
Master of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
Email: clarissa.pah@binus.ac.id

Abstract: This paper explains the use of data mining and fuzzy logic in optimizing decision making for employee recruitment, especially in IT Consultant company. This method addresses the company's need to recruit employees with more objective and accurate by utilizing historical data to find patterns of potential employees for the company. First, classification will be carried out using data mining to determine the predictor attributes of potential employees. These predictor attributes will then be used to arrange fuzzy rule base and fuzzy logic structure which will be used to prioritize employees to be accepted. The results obtained will prioritize employees who are recruited objectively scientifically.

Keywords: Data Mining, Fuzzy Logic, Decision Support, Employee Recruitment

Introduction

Quality workforce in a company is a very important asset to support the company's business goals, both large and small companies (Morgan, 2019). Even, the decision to choose the right employee can predict the success of a job in the future (Guion *et al.*, 2016). Decent employees are an important material in building the success of a company's organization, while unworthy employees will only create many obstacles and spend a lot of resources such as time, energy and money (Khorami and Ehsani, 2015).

An employee selection process will determine the accuracy in choosing qualified employees (Chien and Chen, 2008). The selection process depends on an assessment that distinguishes between prospective employees from one another and predicts the performance of prospective employees in the future (Serhadloğlu *et al.*, 2009). Assessment of employees begins by looking at the predictor attributes attached to employees and compliance with company-determined criteria. The predictor attribute is used to assess the suitability of prospective employees with a position or criteria determined by the company, while the criteria are used to predict the performance of prospective employees (Guion *et al.*, 2016).

Several studies have applied various methods that can automatically assess prospective employees based on the attributes of existing predictors. Some of them use AHP, TOPSIS, fuzzy logic and data mining classification. In this research, researcher combine data mining and fuzzy logic. Data mining classifications can be using to

produce attributes that affect a target. Here, we can separate the class of employees expressed as potential and less potential as targets. Furthermore, researchers can create scenarios that have weight to be able to sort the potential of employees by using the fuzzy logic method.

The contribution of this research is divided into 2, namely towards practitioners and towards research. With practitioners, practitioners more easily determine the prospective employees to be accepted and prospective employees get job positions that fit the practitioners' needs so that they do not waste time due to mismatch positions. For the research/science domain, a model that combines data mining and fuzzy logic is a solution to the gap found in previous studies, namely regarding the selection of attributes according to practical needs with data mining classification methods and more accurate weighting criteria with fuzzy logic.

Fundamentals

This section will explain the case study and theoretical basis for the methods used to resolve the cases.

Personal Selection

Personnel selection plays an important role for controlling the human source and quality in HRM. Effective employee selection is a critical component of a successful organization. Personnel selection is the process of collecting and evaluating information about individuals and choosing those who match the

qualifications needed to perform a predefined job in the best way (Mammadova and Jabrayilova, 2014). Nowadays, organizations face increasingly global competition. Therefore, many academic researchers believe that the personnel selection is a strategic decision because it will impact organizational performance such as customer satisfaction, innovation, quality, profitability and company competitiveness (Khorami and Ehsani, 2015).

Mammadova and Jabrayilova (2014) reviewed the personnel selection studies and found that the several main factors including change in organizations, change in work, change in personnel, change in the society, change of laws and change in marketing have influenced personnel selection. In literature, there are a number of studies which use heuristic methods for employee selection. A fuzzy Multi Criteria Decision Making (MCDM) framework based on the concepts of ideal and anti-ideal solutions for the most appropriate candidate. Also, a fuzzy number ranking method by metric distance for personnel selection problem was proposed and a personnel selection system based on fuzzy AHP was developed in.

Information Technology Consultant (IT) companies can maintain their business through IT project work obtained from their client companies. Most of the employees will be placed in the client's office or commonly referred to on site. Of course, each client company has different characteristics, work habits and pressures, so that not infrequently many employees who when placed, do not feel comfortable with the client company choose to resign from the company, but do not rule out some employees are able to survive in conditions the. Because of the many turnovers that occur, a company that has invested a lot of things for its employees becomes a loss in terms of time for training employees as well as their expenses. Therefore, the accuracy in choosing employees is very important to do and there are many aspects of the prospective employee that must be seen before being accepted so that recruitment is effective and objective.

Fuzzy Logic

Fuzzy Logic is a method that processes parameter values that can be absolute values. In other words, fuzzy logic can process human language that can be a definite value (Utama, 2017).

According to Utama and Taryana (2019) Fuzzy Logic (FL) use is practically broad. This method has been technically employed by many researchers in their research activities as a fundamental method. Basically FL has been used to minimize biased savor when someone adjust a parameter value. Through using common scheme "Fuzzification-De-fuzzification", crisp input is converted to become crisp output before being functioned for further purposes (e.g., prioritizing). Fuzzy logic refers to fuzzy sets that were first published by

Zadeh. Initially, each data or element has one of 2 labels, yes or no (0 or 1). In fuzzy sets, this label is developed with the presence of levels between 0 to 1, meaning that the value of a data or element is no longer 0 or 1 but is elaborated again for example 0.5 which has 50% value 0 and 50% value 1. Calculation of this level can be done with functions standards such as triangles, trapezoidal, Gauss curves depending on the case that want to solve (Karatop *et al.*, 2015).

Mammadova and Jabrayilova (2014) conducted a research on employee hiring decision making by applying the Fuzzy method. According to them, this method can solve the problems of previous studies (multi-criteria assessment and sequencing) that can use qualitative and quantitative data, eliminate the limit on the number of criteria and the number of experts, using a hierarchical structure of criteria. In addition, the key to more objective and transparent decisions is the unlimited number of relevant criteria. The results of this study, fuzzy logic can more accurately determine the selected employees with more specific reasons (because it is adjusted to the tendency of the criteria). But the drawback of this research is that it takes a long time to determine the criteria because it must involve experts. The competence of each expert is different so the criteria and weighting criteria are very dependent on the expertise of the experts. In addition, over time, of course the criteria will be more diverse and may be subject to change according to company needs. Criteria and criteria weighting must always be updated over time. In addition, this fuzzy method is a method that is practical and easy to understand and accurate (with appropriate criteria and weighting) with a wider range of decision levels. Some studies that combine fuzzy logic with other methods such as AHP and TOPSIS produce better and more accurate weighting criteria (Mammadova and Jabrayilova, 2014; 2018).

Data Mining Classification

The data mining classification method uses supervised learning that can classify data according to predetermined label classes. Some algorithms that are often used are decision tree and Naïve Bayes classification. With classification, the model that has been built can predict data samples based on historical data that has been studied previously (Pal and Pal, 2013). The classification method divides the data into several sets, namely training data, validator data and testing data. Training data is used to study data and find patterns that are informative or have important knowledge, data validators have data that are used to validate models that have been formed based on learning outcomes of training data, furthermore and testing data are used to test the accuracy of decision making.

Chien and Chen (2008) have conducted personnel selection using data mining in a high-technology industry. They have compared four classification algorithms of CHAID, CART, ID3 and C4.5. Finally, they were using CHAID for the decision tree because of the most data were categorical values. Furthermore, Azar *et al.* (2013) have done the same approach by using data mining for personnel selection in a commercial bank. They compared four classification algorithms of QUEST, CHAID, C5.0 and CART and found that C5.0 (extension of C4.5) had the best accuracy of 80.43%. Both of Chien and Chen (2008) and Azar *et al.* (2013) using demographic data such as age, sex, marital status, university, degree, mayor, experience year, etc., C4.5 or C5.0 have capability in dealing with numerical and categorical data value at one.

Methodology

The flow diagram shown in Fig. 1 is the steps that will be the reference in this research.

Literature Study

Start with collecting all journal articles, books and other references that explains about decision making in personal selection or employee recruitment. Then, learn about the methods used in the decision support system, especially for employee hiring cases in companies.

Data Mining Process

Cross Industry Standard Process for Data Mining (CRISP-DM) model is used in this research to carry out the data mining process. There are six phase of data mining process as shown at Fig. 2 (Wirth and Hipp, 2000).

Phase 1: Business Understanding

This phase is done by conducting discussion with Human Resource Division manager and staffs (will be called with experts) to understanding the research objectives, requirements from business perspective, determine data mining goals.

Phase 2: Data Understanding

The data understanding will be started with data collection that gained from Human Resource division representative. The data consist of all historical data of employees who have been hired in the IT consultant company. The data are obtained since the employee started to apply a job (demographic data, educational background, working experience, etc.) until employee performance data each year. After that, an attribute definition matrix is produced to ensure that the researchers and the business have same perspective about the data attributes. The matrix consists of attribute name, attribute definition and attribute sample value as shown at Table 1.

Phase 3: Data Preparation

In data preparation, all defined attributes and their values are processed to construct the final dataset that will be used in data mining models. The process starts with attribute selection by choosing only important attributes based on experts' perspectives. Furthermore, data cleaning is applied to eliminate missing values and outliers. Finally, data transformation is applied to simplify the attribute values, for example, the attribute "Average Key Performance Indicator (KIP) Result" has original values A, B, C, D and E, if transformed, from five values can be simplified into two values, namely "Recommended" and "Not recommended", so that the value of A will enter the "Recommended" group, while B, C, D and E will enter the "Not Recommended" group. The objective of data cleansing and transformation is to achieve better quality of data before fed into data mining models (Azar *et al.*, 2013). During data preparation phase, expert representatives are still involved to avoid mistakes.

Next, defining the predictor attributes and target attribute from the dataset (target attribute can be called as label in data mining). Target attribute is set as employee classification whether each hired employee is actually classified as "Recommended" or "Not Recommended". This target attribute is gotten from transformation result of "Average Key Performance Indicator (KIP) Result" original attribute. Meanwhile, the predictor attributes are gotten from all attributes other than label, such as age, marital status, degree, etc., which are used to predict target attribute value.

Phase 4: Modeling

In modeling phase, several selected data mining technique models are applied. The model references can be found when doing literature study on the similar case study documentations. In this case, the models are focused on classification data mining to find out what kind of employee is recommended to be hired or not. The classification model that can be used such as Naïve Bayes, Decision Tree, Random Forest and Support Vector Machine. In this research Rapidminer tool is used to build the data mining models.

Phase 5: Evaluation

In this step, every data mining model are evaluated. Model with highest accuracy is selected as the best classification model. Beside accuracy, classification data mining model also produces important attribute's weight. Attribute weights determine whether certain predictor attributes have a strong effect to the target attribute or not. These attribute weights later calculated to construct rule base matrix on fuzzy logic.

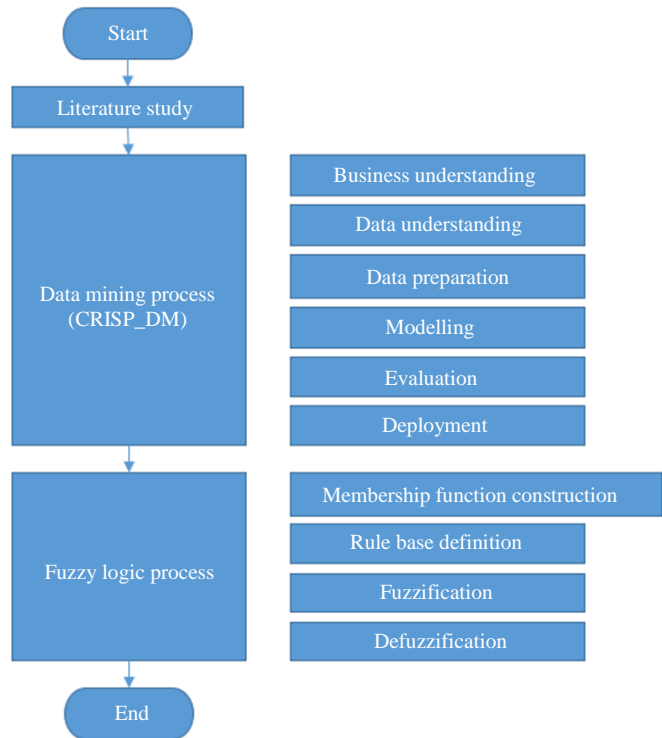


Fig. 1: Research flow for employee selection

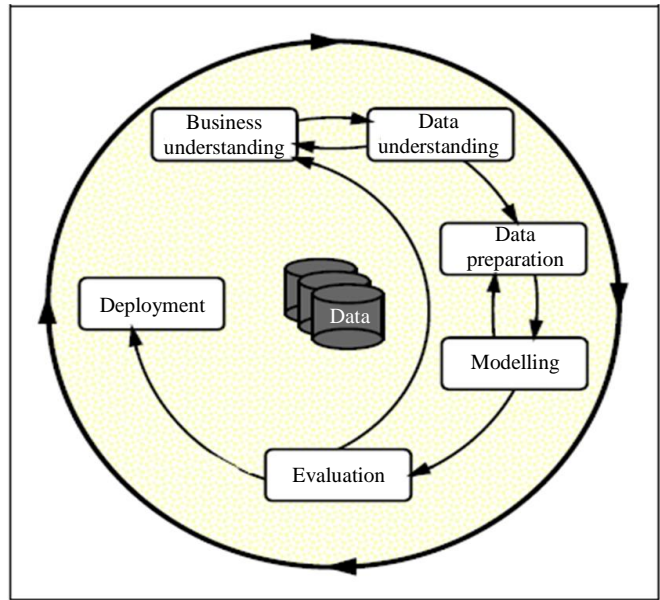


Fig. 2: CRISP-DM process model

Table 1: Example of attribute definition matrix

No	Attribute	Attribute description	Example Value
1	Employee ID	Unique number that identifies the employee and distinguishes one employee from another employee	CEA001, ABC321, DNR122
2	Employee name	The name of employee who has been hired	Adelina Putri
3	Position	Employee position in the company	Consultant, PM, Officer, Sales
4	Role	Employee title/role in the company	Division head, department head
5	Grade	Grade of employee when joining the company	1,2,3,4,5,6,7,8,9,10,11

Phase 6: Deployment

The knowledge that gained from the classification data mining model can be used to develop proper decision support system. In this case, we using the knowledge to build rule base matrix for fuzzy logic process.

Fuzzy Logic Process

Fuzzy logic process in this research is using Mamdani method because of its superiority that is easy to understand and has the ability to produce precise and accurate decisions (Febriany, 2016; Magdalena, 2015).

Step 1: Membership Function Construction

Every attribute value has different importance level (in fuzzy logic can be called Linguistic Variable). The linguistic variable values can be various and depends of the case, but in this case simply uses three values, Not Recommended (NR), Recommended (R) and Highly Recommended (HR). The membership function that used in this case is the combination of trapeze and triangle function as shown in Fig. 3., where the notation for μ_{NR} , μ_R , μ_{HR} is defined as Equation 1 to 3.

$$\mu_{NR} : \begin{cases} 1; X \leq A \\ \frac{B-X}{B-A}; A \leq X \leq B \\ 0; X \geq 10 \end{cases} \quad (1)$$

The notation for μ_R is defined as:

$$\mu_R : \begin{cases} 0; X \leq 0 \text{ OR } X \geq C \\ \frac{X-A}{B-A}; A \leq X \leq B \\ \frac{C-X}{C-B}; B \leq X \leq C \end{cases} \quad (2)$$

The notation for μ_{HR} is defined as:

$$\mu_{HR} : \begin{cases} 0; X \leq B \\ \frac{X-B}{C-B}; B \leq X \leq C \\ 1; X \geq C \end{cases} \quad (3)$$

Step 2: Rule Base Definition

This step is the longest step in fuzzy logic because it should be listing all attribute value combination with the decision. Most studies ask the experts to determine the rules but the more attribute values, the more rules that must be made. So, with help of data mining

attribute results and attribute weights, we can easily construct the rule base.

Rule Base works with combines all attribute value to produces a decision and turns it become rules. The rule can be written as follows (Equation 4):

$$IF (X1 \text{ is } A1 \text{ AND } (X2 \text{ is } A2) \text{ AND } \dots \text{ AND } (Xn \text{ is } An)) \text{ THEN } Y \text{ is } B \quad (4)$$

Where, X is an attribute involved with A as X's attribute value and Y is the decision attribute with B as Y's attribute value. N depends of how many attribute involved in the rule. A is converted from X's attribute value as shown at Table 2. For example, attribute X1 with value X11 is converted to HR as A, so X1 = HR.

To convert each attribute value to linguistic variable, we are using how many percentages of employees with recommended label in the dataset have the attributes value. For example, in Table 3 shows that attribute value X11 has 72% of recommended employees, while X12 only has 28% of recommended employees, so, X11 will has HR as linguistic variable value and X12 has R. This linguistic variable determination is done with the collaboration with experts.

After all attribute values has its linguistic variable, the next step is find the final weight of attribute values with multiply the attribute weight that gained from data mining steps with the recommended employee percentage as shown in Table 4.

From the final weight, the next step is determining the highest total of final weight based on attribute values combination. In Table 4, the highest total is 1.56 that come from combination of X1 = HR AND X2 = HR AND X3 = HR.

Table 2: Attribute value and linguistic variable mapping

Attribute (X)	Attribute value (x)	Linguistic variable (A)
X1	X11	HR
	X12	R
X2	X21	NR
	X22	R
X3	X31	HR
	X32	R
	X32	NR

Table 3: Attribute value conversion to linguistic variable

Attribute (X)	Attribute value (x)	Recommended employees percentages	Linguistic variable (A)
X1	X11	0.72	HR
	X12	0.28	R
X2	X21	0.9	NR
	X22	0.1	R
X3	X31	0.25	HR
	X32	0.50	R
	X33	0.25	NR

Table 4: Final weight of attribute values

Attribute (X)	Attribute Weight (AW)	Attribute Value (x)	Recommended Employees Percentages (REP)	Linguistic Variable (A)	Final Weight (AW* REP)
X ₁	0.86	X ₁₁	0.72	HR	0,62
		X ₁₂	0.28	R	0.24
X ₂	0.75	X ₂₁	0.9	HR	0,68
		X ₂₂	0.1	R	0.08
X ₃	0.51	X ₃₁	0.50	HR	0.26
		X ₃₂	0.37	R	0.19
		X ₃₃	0.13	NR	0,07

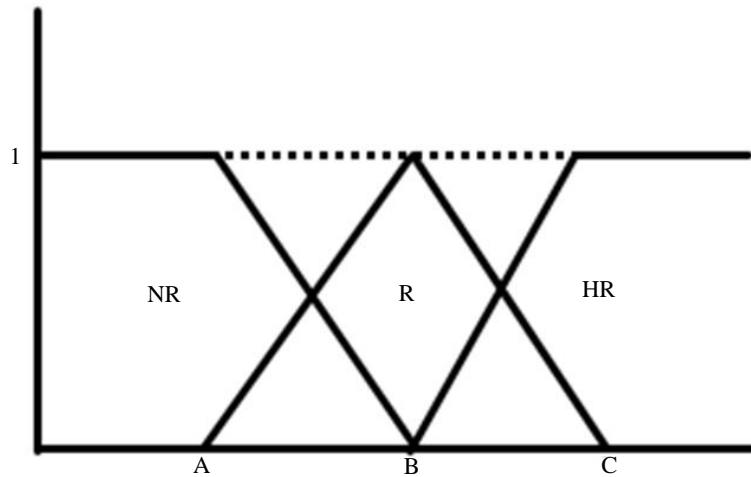


Fig. 3: Membership function example

As result, we can build the rule base matrix with all combination of attribute values and the decision. To find the decisions, we need threshold value for the total final weight each combination. If a combination is equal or greater than the threshold value, then it will has “Yes” as decision, but if the combination is less than the threshold value, the it will has “No” for the decision. In this case, we need expert decision to decide the threshold.

Step 3: Fuzzification

In this research, the fuzzification process will be using Mamdani method. The Mamdani method consists of fuzzy set construction, implication function application, rule composition and defuzzification. Fuzzy set construction is the process to map the input value into the membership function using interpolation linear as suggested by Utama (2017) with following formula (Equation 5):

$$Y = Y_1 + \frac{X - X_1}{X_2 - X_1} (Y_2 - Y_1) \tag{5}$$

Where, X is the input value between X_1 and X_2 value and Y is the μ value between Y_1 and Y_2 (0 and 1). μ value for each X will be used to find implication function (MIN) for all rules with the following formula (Equation 6):

$$\alpha - predicate_i = \min(\mu A_1 [X_1] \cap \dots \cap A_n [X_n]) \tag{6}$$

After calculating implication function for each rule, then the MAX function is applied to make a composition between all rules. This rule composition is using formula (Equation 7):

$$\mu_{sf}(X_i) = \max(\mu_{sf}(X_i), \mu_{kf}(X_i)) \tag{7}$$

Where, $\mu_{sf}(X_i)$ is the μ value of the fuzzy solution until rule- i and $\mu_{kf}(X_i)$ is the μ value of the fuzzy consequent until rule- i . This result is used to construct a new membership function of decision that will be the input of defuzzification process.

Step 4: Defuzzification

This step is done by converts the fuzzy sets that obtained from fuzzification steps into a crisp value. Defuzzification process uses the Centroid method with determine moment, area and center point. To determine the center point, will using the formula (Equation 8):

$$z^* = \frac{\int_a^b \mu A(z) z dz}{\int_a^b \mu A(z) dz} \tag{8}$$

where, z is the result of defuzzification, $\int_a^b \mu A(z)zdz$ is the formula to calculate the moment and $\int_a^b \mu A(z)dz$ to calculate the area. The obtained z value will be used to compare all candidates.

Results and Discussion

To process the data mining until using the fuzzy logic, there will be several steps.

Collecting Data

The collected data in this study are the employees' data of IT Consultant companies which is established from 2007 to 2019 with 247 rows of data with 211 active employees (85.5%) and 36 (14.5%) resigned employees. In this study, only employees' data with "Consultant" position is collected, consists of 170 (68%) with 145 active employees and 25 resigns with existing attributes are as follows: Employee ID, Employee Name, Position, Role, Grade, Joining Date, Length of Service, Sex, Marital Status, Religion, Birth Date, Age, Degree, University, University Accreditation, Employee Status, Resign Flag, Resignation Date, KPI Average, Recruitment Source, Working Experience.

Preprocessing Data

Process data that has been collected previously and determine the attributes that will be used in research. The selected attribute is an attribute that has the possibility of influencing employee selection.

Feature Selection: From all the attributes, only 7 attributes are taken as predictor attributes and 1 attribute as labels:

- Target attribute/Label: Performance Assessment. The attribute has 2 values: "Yes" refers to recommended employees and "No" refers to not recommended employees. This attribute is come from KIP Average with Outstanding value for "Yes" and other than that will marked "No". In

membership function, this attribute will be called Decision attribute

- Predictor Attributes: Working Experience (number of years), Sex (male, female), Marital Status (single, married), Join Age (difference between join date and birthdate), Degree (S1, Under S1, S2), Recruitment Source (Internal, External), University Accreditation (A, B, C)

The selected features/attributes are gained based on related works from the previous research and experts' perspective. Table 5 shows the related works' predictor attributes.

Data Cleansing: Data cleansing is done when there is inconsistent or duplicated data. The steps taken are fill in the missing value with the mean or mode of the attribute value and deleting redundant data manually. For the data used in this study, the data obtained is no redundant.

Data Transformation: Data transformation is conducted to do data normalization, eliminate outliers and transform attributes value.

Data Mining Classification

Data mining experiment which has conducted in this research were using Rapidminer Studio tool.

Figure 4 shows the process of decision tree using C4.5 using Rapidminer tool.

Several classification methods are tried to get the most accurate classification, such as Decision Tree (C4.5), Random Forest, Support Vector Machine (SVM) and Naïve Bayes.

The results obtained in Fig. 5 state that the decision tree has the highest accuracy of 87.87%, followed by SVM, Random Forest and Naïve Bayes.

Based on the accuracy, decision tree (C4.5) technique is used to make rules of classification and to calculate the attribute weight of importance. Fig. 6 shows the result of decision tree for recruitment classification. It shows that only 6 attributes have used to make a classification tree.

Table 5: Related works' predictor attributes

No	Related works' predictor attributes	References
1	Age	Chien and Chen (2008; Azar <i>et al.</i> , 2013; Supriaty <i>et al.</i> , 2016; Harpad, 2018; Tai and Hsu, 2006)
2	Gender	Chien and Chen (2008; Azar <i>et al.</i> , 2013)
3	Marital Status	Chien and Chen (2008; Azar <i>et al.</i> , 2013)
4	Degree	Chien and Chen (2008; Azar <i>et al.</i> , 2013; Lytvyn <i>et al.</i> , 2016; Mammadova and Jabrayilova, 2018; Harpad, 2018; Tai and Hsu, 2006)
5	University/School	Chien and Chen (2008; Kelemenis and Askounis, 2010; Azar <i>et al.</i> , 2013; Lytvyn <i>et al.</i> , 2016)
6	University Majority	Chien and Chen (2008; Azar <i>et al.</i> , 2013)
7	Work Experience	Chien and Chen (2008; Azar <i>et al.</i> , 2013; Karatop <i>et al.</i> , 2015; Supriaty <i>et al.</i> , 2016; Tai and Hsu, 2006)
8	Recruitment Source	Chien and Chen (2008)

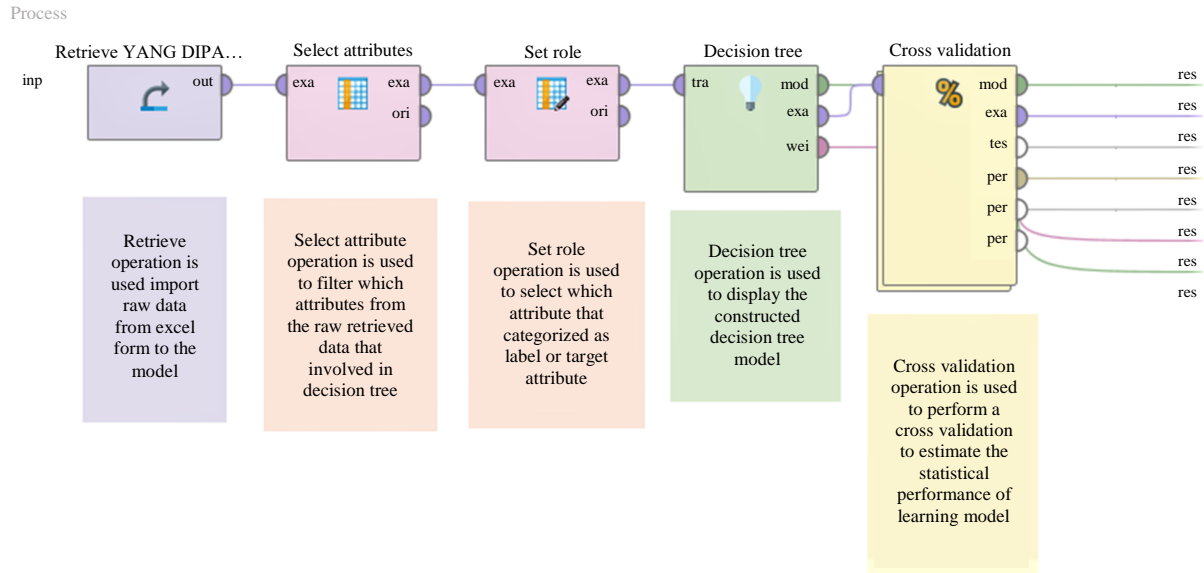


Fig. 4: C.4.5 Decision tree process using rapidminer

Model	Accuracy	Standard de...	Total time	Training time (1,0...
Naïve Bayes	83.1%	±6.4%	138 ms	~ 0 ms
Decision tree	87.8%	±4.4%	195 ms	~ 0 ms
Random forest	87.8%	±4.4%	2 ms	49 ms
Support vector machine	87.8%	±4.4%	822 ms	10 ms

Fig. 5: Data mining accuracy comparison

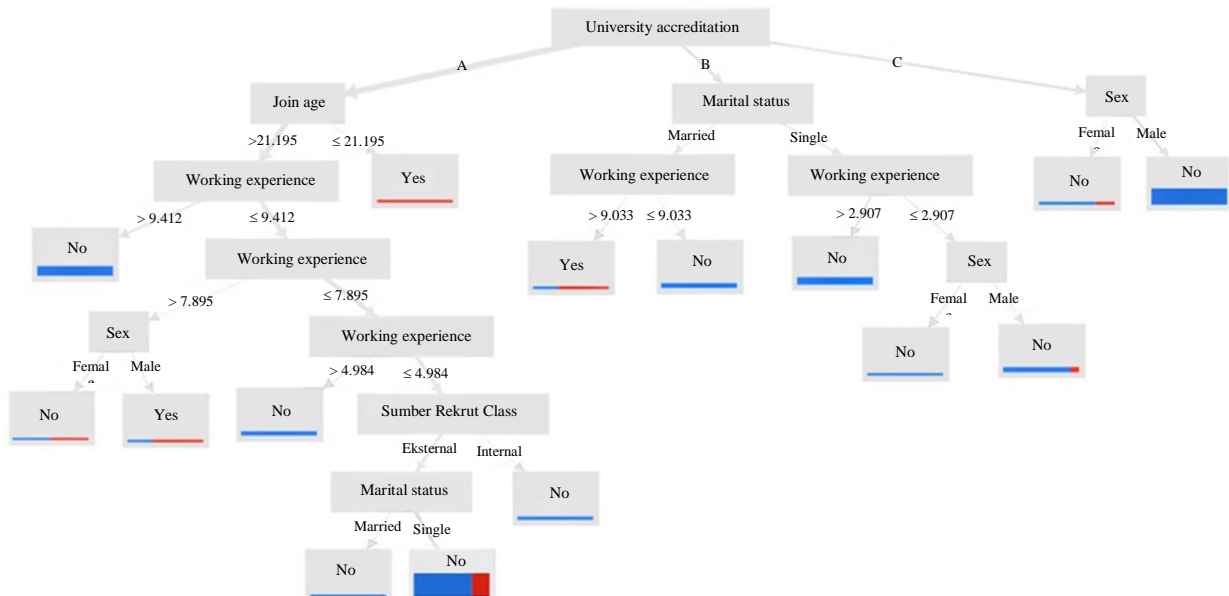


Fig. 6: C.4.5 decision tree for employee recruitment

The tree explains that:

University Accreditation = A
 | Join Age > 21.195
 || Working Experience > 9.412: NO {NO = 19, YES = 0}
 || Working Experience ≤ 9.412
 ||| Working Experience > 7.895
 |||| Sex = Female: NO {NO = 1, YES = 1}
 |||| Sex = Male: YES {NO = 1, YES = 2}
 ||| Working Experience ≤ 7.895
 |||| Working Experience > 4.984: NO {NO = 7, YES = 0}
 |||| Working Experience ≤ 4.984
 ||||| Sumber Rekrut Class = External
 ||||| Marital Status = Married: NO {NO = 4, YES = 0}
 ||||| Marital Status = Single: NO {NO = 40, YES = 12}
 ||||| Sumber Rekrut Class = Internal: NO {NO = 5, YES = 0}
 | Join Age ≤ 21.195: YES {NO = 0, YES = 3}
 University Accreditation = B
 | Marital Status = Married
 || Working Experience > 9.033: YES {NO = 1, YES = 2}
 || Working Experience ≤ 9.033: NO {NO = 8, YES = 0}
 | Marital Status = Single
 || Working Experience > 2.907: NO {NO = 14, YES = 0}
 || Working Experience ≤ 2.907
 ||| Sex = Female: NO {NO = 3, YES = 0}
 ||| Sex = Male: NO {NO = 8, YES = 1}
 University Accreditation = C
 | Sex = Female: NO {NO = 3, YES = 1}
 | Sex = Male: NO {NO = 34, YES = 0}

From the related works, Table 6 shows the mapping of related works' attributes with the experts' attributes.

Even though the accuracy of the data mining model is high, not necessarily the precision level is also high because there is imbalance between the labels "Yes" and "No". This is proved by the precisions produced, where "No" prediction gains 88.82% precision and "Yes" prediction only gains 44.44% precision. Therefore, to make it more precise, the expert perspectives will be involved by combining the experts' attribute weight that obtained from attribute rating surveys and the data mining's attribute weight of importance.

Table 7 shows the average weight of the attributes which used in data mining and from the experts' perspective. This weight will be used to fuzzy logic process to recalculate the decision to assess the personnel candidate whether accepted or rejected.

Fuzzy Logic Implementation

The fuzzy logic process starts with determine the membership function of each involved attribute. They are:

Figure 7 shows the membership function of Join Age attribute where (Equitation 9 to 11):

$$\mu_{JoinAgeHR} = \begin{cases} \frac{23-X}{23-22}; 22 \leq X \leq 23 \\ 0; X \geq 23 \end{cases} \quad (9)$$

$$\mu_{JoinAgeR} = \begin{cases} 0; X \leq 22 \text{ OR } X \geq 24 \\ \frac{X-22}{23-22}; 22 \leq X \leq 23 \\ \frac{24-X}{24-23}; 23 \leq X \leq 24 \end{cases} \quad (10)$$

$$\mu_{JoinAgeNR} = \begin{cases} 0; X \leq 23 \\ \frac{X-23}{24-23}; 23 \leq X \leq 24 \\ 1; X \geq 24 \end{cases} \quad (11)$$

Figure 8 shows the membership function of University Accreditation attribute where (Equitation 12 to 14):

$$\mu_{UniversityAccreditationNR} = \begin{cases} 1; X \leq 2.5 \\ \frac{5-X}{5-2.5}; 2.5 \leq X \leq 5 \\ 0; X \geq 5 \end{cases} \quad (12)$$

$$\mu_{UniversityAccreditationR} = \begin{cases} 0; X \leq 2.5 \text{ OR } X \geq 7.5 \\ \frac{X-2.5}{5-2.5}; 2.5 \leq X \leq 5 \\ \frac{7.5-X}{7.5-5}; 5 \leq X \leq 7.5 \end{cases} \quad (13)$$

$$\mu_{UniversityAccreditationHR} = \begin{cases} 0; X \leq 5 \\ \frac{X-5}{7.5-5}; 5 \leq X \leq 7.5 \\ 1; X \geq 7.5 \end{cases} \quad (14)$$

Figure 9 shows the membership function of Degree attribute where (Equitation 15 to 17):

$$\mu_{DegreeNR} = \begin{cases} 1; X \leq 2.5 \\ \frac{5-X}{5-2.5}; 2.5 \leq X \leq 5 \\ 0; X \geq 5 \end{cases} \quad (15)$$

$$\mu_{DegreeR} = \begin{cases} 0; X \leq 2.5 \text{ OR } X \geq 7.5 \\ \frac{X-2.5}{5-2.5}; 2.5 \leq X \leq 5 \\ \frac{7.5-X}{7.5-5}; 5 \leq X \leq 7.5 \end{cases} \quad (16)$$

$$\mu_{DegreeHR} = \begin{cases} 0; X \leq 5 \\ \frac{X-5}{7.5-5}; 5 \leq X \leq 7.5 \\ 1; X \geq 7.5 \end{cases} \quad (17)$$

Figure 10 shows the membership function of Marital Status attribute where (Equitation 18 and 19):

$$\mu_{MaritalStatusNR} = \begin{cases} 1; X \leq 5 \\ \frac{7.5-X}{7.5-5}; 5 \leq X \leq 7.5 \\ 0; X \geq 7.5 \end{cases} \quad (18)$$

$$\mu_{MaritalStatusR} = \begin{cases} 0; X \leq 5 \\ \frac{X-5}{7.5-5}; 5 \leq X \leq 7.5 \\ 1; X \geq 7.5 \end{cases} \quad (19)$$

Figure 11 shows the membership function of Recruitment Source attribute where (Equitation 20 and 21):

$$\mu_{RecruitmentSourceNR} = \begin{cases} 1; X \leq 5 \\ \frac{7.5-X}{7.5-5}; 5 \leq X \leq 7.5 \\ 0; X \geq 7.5 \end{cases} \quad (20)$$

$$\mu_{RecruitmentSourceR} = \begin{cases} 0; X \leq 5 \\ \frac{X-5}{7.5-5}; 5 \leq X \leq 7.5 \\ 1; X \geq 7.5 \end{cases} \quad (21)$$

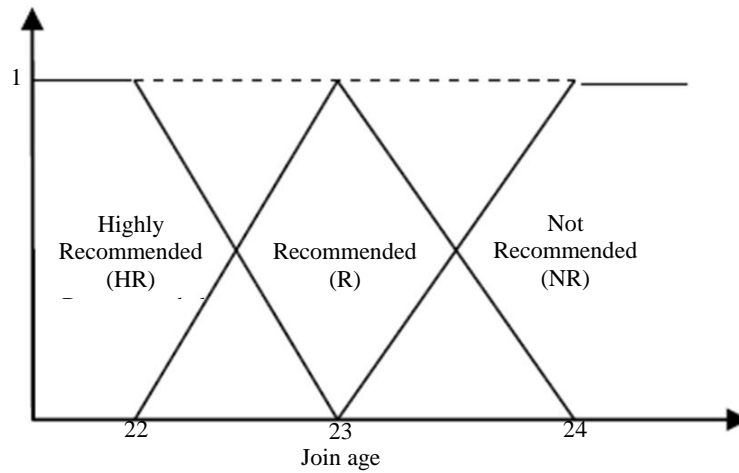


Fig. 7: Join age membership function

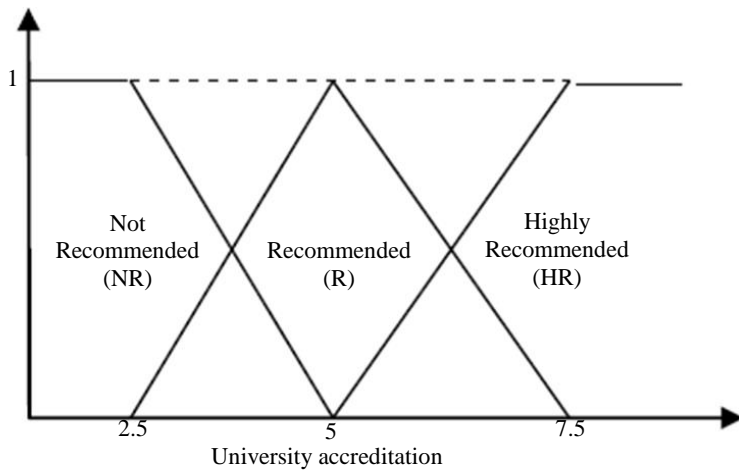


Fig. 8: University accreditation membership function

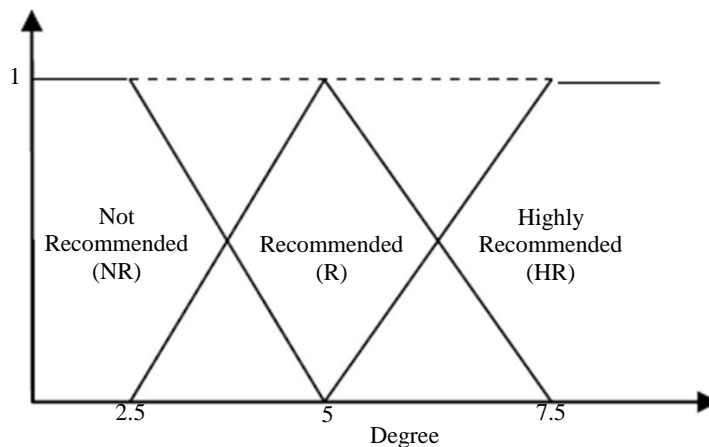


Fig. 9: Degree membership function

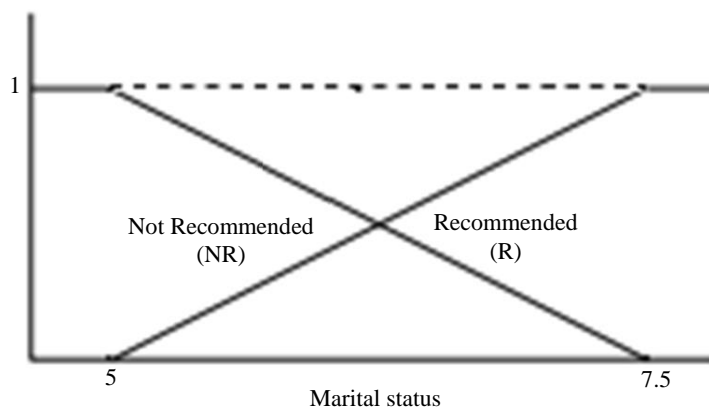


Fig. 10: Marital status membership function

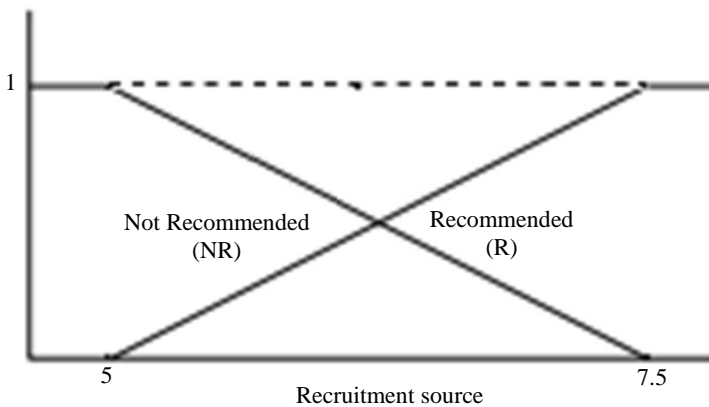


Fig. 11: Recruitment source membership function

Table 6: Predictor attributes of related works and experts mapping

No	Predictor attributes	Experts	Related works
1	Join age	V	V
2	Working experience	V	V
3	Degree class	V	V
4	University accreditation	-	V
5	Recruitment source	V	V
6	Marital status	V	V
7	Sex	V	V

Table 7: Average weight of attributes

No	Attributes	Data mining weight (A)	Experts weight (B)	Average weight (A+B)/2
1	WE	0.63	0.76	0.70
2	JA	0.08	0.74	0.41
3	DE	-	0.64	0.32
4	UA	0.04	0.51	0.28
5	SE	0.15	0.33	0.24
6	RS	0.03	0.36	0.20
7	MS	0.07	0.30	0.18

Figure 12 shows the membership function of Sex attribute where (Equitation 22 and 23):

$$\mu_{SexR} = \begin{cases} 1; X \leq 5 \\ \frac{7.5 - X}{7.5 - 5}; 5 \leq X \leq 7.5 \\ 0; X \geq 7.5 \end{cases} \quad (22)$$

$$\mu_{SexHR} = \begin{cases} 0; X \leq 5 \\ \frac{X - 5}{7.5 - 5}; 5 \leq X \leq 7.5 \\ 1; X \geq 7.5 \end{cases} \quad (23)$$

Figure 13 shows the membership function of Working Experience attribute where (Equitation 24 and 25):

$$\mu_{WorkingExperienceR} = \begin{cases} 1; X \leq 2 \\ \frac{10 - X}{10 - 2}; 2 \leq X \leq 10 \\ 0; X \geq 10 \end{cases} \quad (24)$$

$$\mu_{WorkingExperienceHR} = \begin{cases} 0; X \leq 2 \\ \frac{X - 2}{10 - 2}; 2 \leq X \leq 10 \\ 1; X \geq 10 \end{cases} \quad (25)$$

Figure 14 shows the membership function of Decision attribute where (Equitation 26 and 27):

$$\mu_{DecisionNO} = \begin{cases} 1; X \leq 5 \\ \frac{7.5 - X}{7.5 - 5}; 5 \leq X \leq 7.5 \\ 0; X \geq 7.5 \end{cases} \quad (26)$$

$$\mu_{DecisionYES} = \begin{cases} 0; X \leq 5 \\ \frac{X - 5}{7.5 - 5}; 5 \leq X \leq 7.5 \\ 1; X \geq 7.5 \end{cases} \quad (27)$$

After defining all membership functions, the next step is defining the rule base matrix as shown in Table 10. There are 432 rules with 110 rules are decided as “Yes” and 332 rules are decided as “No”. The decision is calculated with the threshold value (Equitation 28):

$$\text{Threshold Value} = 0.75 * \text{Highest rule weight} \quad (28)$$

Where, 0.75 is come from experts’ perspective and the highest rule weight is come from total attributes’ weight from Equitation 29. So the Highest Rule Weight gained is 0.67 and the Threshold value gained is 0.5. If a rule weight is more the highest rule weight is come from (Equitation 29):

$$\begin{aligned} & \text{IF Recruitment Source} = R \text{ AND } S_{ex} = HR \\ & \text{AND Marital Status} = \text{Single AND University} \\ & \text{Accreditation} = HR \text{ AND Working Experience} \\ & = HR \text{ AND Join Age} = HR \text{ AND Degree Class} = HR \end{aligned} \quad (29)$$

And the 0.75 is come from experts’ perspective that distinguish fulfil 75% of highest rule weight (100%) and vice versa. which rules can.

Furthermore, fuzzification process is done with entering the input values. The input values are mapped into the defined membership function. For the categorical values, we ask the experts decision to determine the input value in numerical format. For example, there are following inputs.

- Join Age: 24 years’ old with X = 24
- Working Experience: 3 years with X = 3
- University Accreditation: A with X = 6.5
- Marital Status: Single with X = 7
- Degree: S1 with X = 7
- Rec Source: External with X = 7
- Sex: Male with X = 6

By using linear interpolation of the membership function, we got membership function attribute value as shown in Table 8.

The gained membership function attribute values are combined using implication function (MIN). Here are the examples:

[R115] IF Join Age is Not Recommended AND Working Experience is Recommended AND University Accreditation is Recommended AND Marital Status is Not Recommended AND Degree is Not Recommended AND Recruitment Source is Recommended AND Sex is Recommended, THEN Decision is NO:

$$\alpha - \text{predicate115} = \min(0, 0.125, 0.4, 0.2, 0, 0.8, 0.6) = 0$$

[R12] F Join Age is Highly Recommended AND Working Experience is Highly Recommended AND University Accreditation is Recommended AND Marital Status is Not Recommended AND Degree is Not Recommended AND Recruitment Source is Not Recommended AND Sex is Recommended, THEN Decision is YES:

$$\alpha - \text{predicate}_{12} = \min(0, 0.875, 0.4, 0.2, 0.2, 0.6) = 0$$

[R52] IF Join Age is Not Recommended AND Working Experience is Highly Recommended AND University Accreditation is Highly Recommended AND Marital Status is Not Recommended AND Degree is Recommended AND Recruitment Source is Not Recommended AND Sex is Recommended, THEN Decision is NO:

$$\alpha - \text{predicate}_{52} = \min(1, 0.875, 0.6, 0.2, 0.2, 0.2, 0.6) = 0.2$$

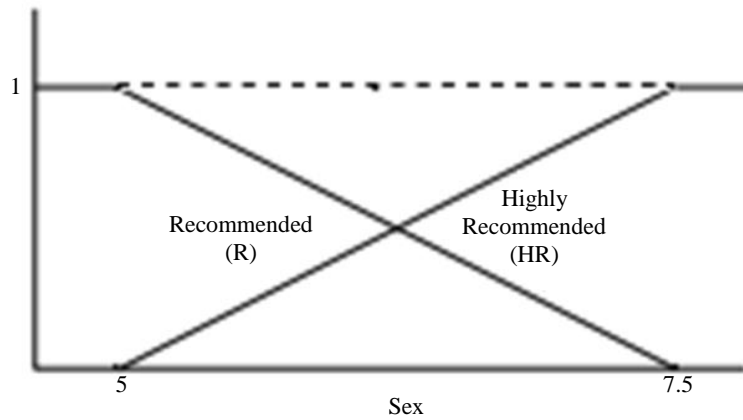


Fig. 12: Sex membership function

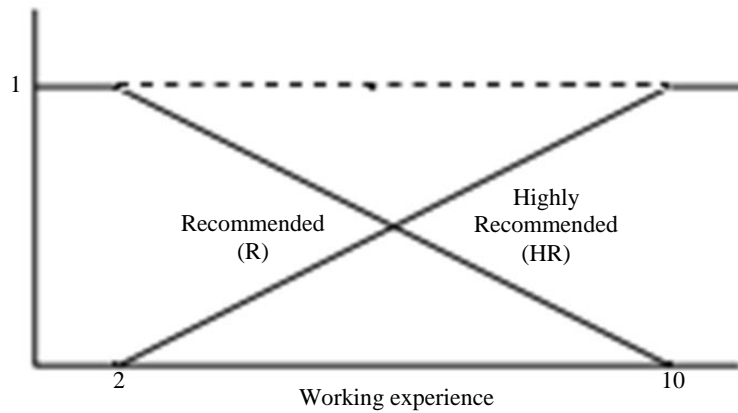


Fig. 13: Working experience membership function

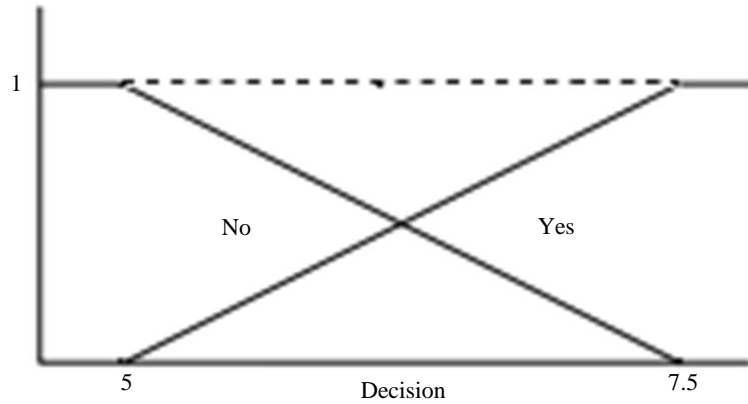


Fig. 14: Decision membership function

Table 8: Membership function attributes and values

Membership function attribute	Membership function attribute value
$\mu_{JoinAgeNR}$	1
$\mu_{JoinAgeR}$	0
$\mu_{JoinAgeHR}$	0
$\mu_{WorkingExperienceR}$	0.125
$\mu_{WorkingExperienceHR}$	0.875
$\mu_{UniversityAccreditationNR}$	0
$\mu_{UniversityAccreditationR}$	0.4
$\mu_{UniversityAccreditationHR}$	0.6
$\mu_{MaritalStatusNR}$	0.2
$\mu_{MaritalStatusR}$	0.8
$\mu_{DegreeNR}$	0
$\mu_{DegreeR}$	0.2
$\mu_{DegreeHR}$	0.8
$\mu_{RecruitmentSourceNR}$	0.2
$\mu_{RecruitmentSourceR}$	0.8
μ_{SexR}	0.6
μ_{SexHR}	0.6

Table 9: Moment and area values

Area name	Moment values	Area values
D1	10.8	3.6
D2	2.9	0.45
D3	0	0

Table 10: Example of rule base matrix

Rule	Input							Value							Output Decision	
	JA	WE	UA	MS	DE	RS	SE	JA	WE	UA	MS	DE	RS	SE		SUM
1	NR	R	NR	NR	NR	NR	R	0.06	0.07	0.01	0.02	-	-	0.04	0.20	NO
2	R	R	NR	NR	NR	NR	R	0.14	0.07	0.01	0.02	-	-	0.04	0.27	NO
3	HR	R	NR	NR	NR	NR	R	0.55	0.07	0.01	0.02	-	-	0.04	0.69	NO
4	NR	HR	NR	NR	NR	NR	R	0.06	0.18	0.01	0.02	-	-	0.04	0.31	NO
5	HR	HR	R	NR	NR	NR	R	0.55	0.18	0.04	0.02	-	-	0.04	0.83	YES

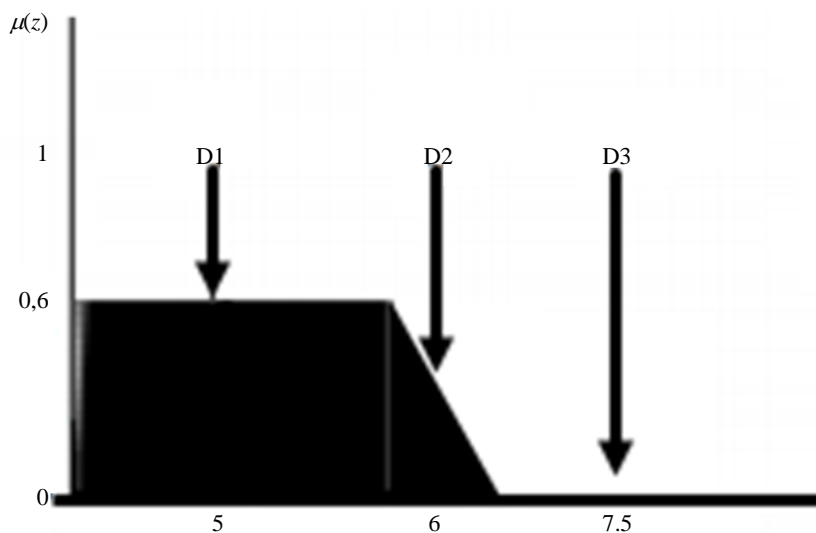


Fig. 15: Defuzzification membership function

After calculating the application of the implication function for each rule, then the MAX function is used to do the composition between all the rules and gains:

$$\begin{aligned} \mu_{Composition}(X) &= \\ \max(\mu_{DecisionNO}(X), \mu_{DecisionYES}(X)) &= \\ \max(0.6, 0) & \end{aligned} \quad (30)$$

Where, these values are then entered into the (26) and (27) formulas and get the result of $Z = 6$ for $\mu_{DecisionNO}(Z) = 0.6$ and $Z = 7.5$ for $\mu_{DecisionYES}(Z) = 0$.

As result, a new membership function is formed and will be used for defuzzification step (Equitation 31):

$$\mu(Z) = \begin{cases} 0.6; Z \leq 6 \\ \frac{7.5-Z}{7.5-6}; 6 \leq Z \leq 7.5 \\ 0; Z \geq 7.5 \end{cases} \quad (31)$$

Figure 15 shows the area that are formed by the defined membership function in Equation (31).

This area is divided in 3 area, D1, D2 and D3.

So, using the formula from Equation (8), the moment and the area values are defined in Table 9 and the result of Z is 3.388.

Based on the Equitation (26) and (27), if Z is under 5, the candidate will be automatically categorized as "No". So in this case, the candidate is rejected.

Conclusion

This paper has explained the detail of data mining process to classify the prospective employee whether will be accepted or rejected based on experimental or historical data. Researcher found that C.4.5 data mining model is the most accurate model to be used to find involved attributes with their importance weight. The data mining is indeed a very good way to avoid the subjectivity in decision making, but in this case study, the collected data tends to be small and even though with high accuracy, the precision of classification is still low because of imbalance classes of label (total of "Yes" and "No" in classification). Therefore, fuzzy logic is used to increase the validity of decision by collaboration with experts in attribute weighting. And finally the methods build can produce decision with specific weight which can distinguish between one prospective employee and others prospective employee.

The scope of this research is limited by total historical employee data and the used attributes. In future research, will be better if using bigger historical data and add more attributes such as IQ, EQ, Personality test result, English proficiency test and other related test to optimize the produced method later.

Acknowledgement

We would like to thank Bina Nusantara University who has supported and sponsored our studies and works, particularly Bina Nusantara Graduate Program, Master of Computer Science.

Author's Contribution

Clarissa Elfira Amos Pah: Has contributed to the article writing, methodology analysis and simulation of the operated methodology.

Ditdit Nugeraha Utama: Has contributed to supervising and directing the operated methodologies and paper writing.

Ethics

This article is genuine and has never been published. All involved authors have reviewed and agreed to submit this article with no ethical issue.

References

- Azar, A., M. Sebt, P. Ahmadi and A. Rajaeian, 2013. A model for personnel selection with a data mining approach: A case study in a commercial bank. SA J. Human Resource Manage.
- Chien, C.F. and L.F. Chen, 2008. Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. Exp. Syst. Applic., 34: 280-290. DOI: 10.1016/j.eswa.2006.09.003
- Febriany, N., 2016. Aplikasi metode fuzzy mamdani dalam penentuan status gizi dan kebutuhan kalori harian balita menggunakan software MATLAB. J. EurekaMatika, 5: 84-96.
- Guion, R.M., S. Highhouse and D. Doverspike, 2016. Essentials of Personnel Assessment and Selection. 2nd Edn., Routledge, ISBN-10: 1317427793, pp: 286.
- Harpad, B., 2018. Penerapan metode Analytical Hierarchy Process (AHP) pada pemilihan staf laboratorium komputer STMIK widya cipta dharma Samarinda. SEBATIK STMIK WICIDA.
- Karatop, B., C. Kubat and O. Uygun, 2015. Talent management in manufacturing system using fuzzy logic approach. Comput. Industrial Eng., 86: 127-136. DOI: 10.1016/j.cie.2014.09.015
- Kelemenis, A. and D. Askounis, 2010. A new TOPSIS-based multi-criteria approach to personel selection. Exp. Syst. Applic., 37: 4999-5008. DOI: 10.1016/j.eswa.2009.12.013
- Khorami, M. and R. Ehsani, 2015. Application of multi criteria decision making approaches for personnel selection problem: A survey. Mahdi Khorami Int. J. Eng. Res. Applic., 5: 14-29.

- Lytvyn, V., V. Vysotska, P. Pukach, I. Bobyk and B. Pakholok, 2016. A method for constructing recruitment rules based on the analysis of a specialist's competences. *Eastern-Eur. J. Enterprise Technol.*, 6: 4-16.
DOI: 10.15587/1729-4061.2016.85454
- Magdalena, L., 2015. Fuzzy Rule-Based System. In: *Evolutionary Computation and Constraint Satisfaction*, Kacprzyk, J. and W. Pedrycz (Eds.), Springer, pp: 203-218.
- Mammadova, M. and Z. Jabrayilova, 2014. Application of fuzzy optimization method in decision-making for personnel selection. *Intell. Control Automation*, 5: 190-204. DOI: 10.4236/ica.2014.54021
- Mammadova, M.H. and Z.G. Jabrayilova, 2018. Decision-making support in human resource management based on multi-objective optimization. *TWMS J. Applied Eng. Math.*, 9: 55-72.
- Morgan, D.B., 2019. Management strategies for reducing voluntary employee turnover in small. PhD Thesis, Business Administration, Walden University, Minneapolis, Hennepin.
- Pal, A.K. and S. Pal, 2013. Evaluation of teacher's performance: A data mining approach. *Int. J. Comput. Sci. Mobile Comput.*, 2: 359-369.
DOI: 10.4236/ica.2014.54021
- Serhadloğlu, G., Z. Güngöra and S.E. Kesen, 2009. A fuzzy AHP approach to personnel selection problem. *Applied Soft Comput.*, 9: 641-646. DOI: 10.1016/j.asoc.2008.09.003
- Supriaty, B., R. Malani and O.D. Nurhayati, 2016. Design of information system for acceptance selection of prospective employees online using tahani fuzzy logic method and Simple Additive Weighting (SAW). *Int. J. Comput. Inform.*, 1: 17-28. DOI: 10.19732/10.19732/vol1122016
- Tai, W.S. and C.C. Hsu, 2006. A realistic personnel selection tool based on fuzzy data mining method.
- Utama, D. and U. Taryana, 2019. Fuzzy logic for simply prioritizing information in academic information system. *Int. J. Mech. Eng. Technol.*, 10: 1594-1602.
- Utama, D.N., 2017. *Sistem Penunjang Keputusan: Filosofi, Teori dan Implementasi*. 1st Edn., Penerbit Garudhawaca, Yogyakarta, ISBN-13: 9786026581013, pp: 123.
- Wirth, R. and J. Hipp, 2000. *CRISP-DM: Towards a standard process model for data mining*.