Particle Swarm Optimization Based C4.5 for Teacher Performance Classification

Abdul Rahman Kadafi[1] and Windu Gata[2]
{abdurrahman.kadafi@gmail.com[1], windu.gata@gmail.com[2]}

Post Graduate Master of Computer Science,College of Information and Computer Management Nusa Mandiri Kramat Raya Road No. 18 Central Jakarta, Indonesia

Abstract: The role of teachers on management of learning to meet the standards of determined competence, is one of key to success. An integrated Islāmic school has a unique standard of teacher competence, there are soft competencies and hard competencies. This research has to get better of classification algorithm to find the pattern of linkage between soft competencies and teacher performance assessed based on hard competencies aspect. The algorithm used is C4.5 algorithm with attribute selection using particle swarm optimization (PSO). The results showed that C4.5 algorithm with PSO resulted accuracy 79.70% with kappa value of 0.528 and 0.19 (enough class). The soft competencies attribute has the highest influence on teacher performance is achievement orientation. From the result of accuracy and kappa, can be concluded that in classification data of teacher for performance aspect and soft competencies. The dominant factor that influence teacher performance base on soft competencies is achievement orientation, then conseptual thinking and concenrn for order, communication and organizational commitment.

Keywords: Data Mining, C4.5, PSO, Soft Competencies

1. Introduction

The role of teachers on management learning is the success way to support the student achievement. The successful of implementation 2013 curriculum in school are can't be escape by the success teacher in managing of learning. The demans of curriculum wants a learning that oriented to student (student oriented), scientific thematic learning demands the competence of teacher to manage the learning optimally (Sri, 2015). Teachers can be the one of spearhead succes of education in school. As good as curriculum applied, if the teacher doesn't have a good competence to do the curriculum in learning process at class, achieving student competencies will be difficult to improve. Achievement of educational institutions can be seen from the achievement of teachers (Sanjay & Keshav, n.d.) (Asanbe and William, 2016).

Intregated Islamic School have a typical standards of education personel if comparing with the other schools. The standard of teacher competence have two things, there are soft competencies and hard competencies. There a ten of soft competencies of the basic competence that the teachers need to have as educators, that would be the research attribute in this study (Fahmy, 2010). Soft skill be important things to discuss because the success of human majority affected by their soft skills when facing the challenge. The ability of soft skill is the essential competence that important to developed up in early on as foundation and provision on education and in life next. With a balanced of hard and soft competence, the childrens will be more competitive.

The children will reach the success by the adequate soft skill. It's same like the research from Harvard University said a succes of someone career 80% are determined by soft skill then 20% are determined by hard skill (Sri, 2015). The competency model is important to do, because the formation of competency modeling are identified as needed from a finding through competency exploration which is related by organizational aspirations and be the determinant of succes in doing the work and it is evident in Japan Industry (Rizal et al., 2013).

Data mining can help predict conditions based on available past data (Gupta, Kumar, & Sharma, 2011). The resultant tree (C4.5) is a existing conventional classifiers for classification Problem (Tsai, Chen, Su, & Lin, 2012). Decision Tree Algorithm be the better algorithm than the other algorithm which have the precise accuracy (Nugroho, Witarto, & Handoko, 2003). C4.5 algorihtm have a power to create a model to predict class labels with new data or with previously unknown data well, and have a speed and efficiency computing time that needed to create and using the models (Charru, 2015). C4.5 algorihtm can give the recommendation model of employee performance valuation (Abadi, 2016).

PSO is the smart random search algorithm (Galway, 2011). Particle swarm optimization (PSO) is a method that can be used to improve a solution with a certain quality measure (Garikapati, Penki, & Gogineni, 2017). In PSO technique have some way to do optimization there are to increase attribute quality (attribute weight) to all of attribute or variable used, selecting attribute (attribute selection) and feature selection. PSO is a very simple optimization technique to apply and modifiy some parameter (Samad, Basari, Hussin, Pramudya, & Zeniarja, 2013). Seelection techniqueques feature of PSO in research is using as technique to increase the accuracy of C4.5 algorihtm.

The aim of research is to know the level accuracy of C4.5 algorihtm in the data classification after the optimization by using the PSO as feature selection for teacher soft competencies.

2. Research Method

At the stages of research process data mining, research design method using CRISP-DM (Cross-Industry Standard Process for Data Mining) as all processes, preprocessing data, model formation, evaluation model, and last deployment (Larose, 2005). In this research execution, the research design method which refers to CRISP-DM method is in the follows figure:
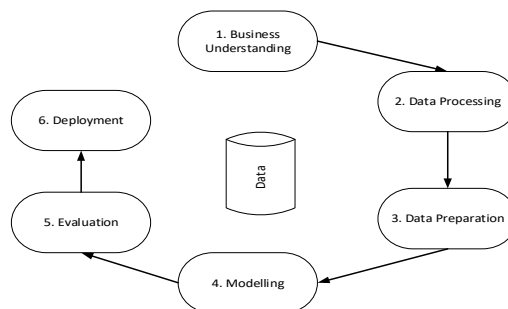

Fig. 1. Research Step.

Research stages are based on CRISP-DM model, with the following stages:
1. Business Understanding, defined about the knowledge that acquired in form of the general questions.

2. Data Understanding, to collect, identifiy, and understand the data assets owned. Desribes the data source and attributes of data used.
3. Data Preparation, includes of data cleansing activities,reformatting the data,reduce the amount of data and etc to prepare data to be consistent in required format.
4. Modeling, trial and implementation of the C4.5 algorihtm to get the model as computational representation from observation result which is the result of search and identification of pattrens contained in data.

Algoritm C4.5 have the same basic principals of work with algorithm ID3. The main difference of C4.5 from ID3 are :
1. C4.5 can handle the continuous and discrete attribute.
2. C4.5 can handle the traning dta with missing value.
3. The result of decision tree C4.5 will be trimmed after formated.
4. The choosing attribute using Gain Ratio. Gain information in ID3 is prioritized in the test that produced many outputs.

To counting the gain ratio in C4.5 algorithm, used this following formula (Xindong And Vipin, 2009) :

$$\text{Entrophy (s)} = \sum_{i=1}^{n} - \text{pi Log 2 pi} \qquad (1)$$

Explanation :
- S is the set (dataset) of case
- n is number of partition S
- pi is the probability from sum (Ya) divided by total cases

$$\text{Gain(s)} = \text{Enttrophy (s)} - \sum_{k=1}^{n} \frac{Si}{S} * \text{Entrophy (si)} \qquad (2)$$

$$\text{Gain Ratio (S, A)} = \frac{\text{Gain(S, A)}}{\text{SplitInfo(S, A)}} \qquad (3)$$

To calculate the value of splitinfo that states as potential information used by the following:

$$\text{Splitinfo(s)} = - \sum_{k=1}^{n} \frac{Si}{S} \text{Log 2} \frac{Si}{s} \qquad (4)$$

The reason why use the gain ratio (a) in C4.5 (not gain (a)) as criterion in selection the attribute is the gain biased to attribute that have many unique values.

PSO algorithm using two random variables r1 and r2 which both produce random numbers with spacing between 0 and 1. Random variables are used to give effect to stochastic character of PSO algorithm. The values of r1 and r2 are adjusted with constants c1 dan c2

which have a range of value between $0 < c1, c2 \leq 2$. The constants is called as acceleration of cefficient that affect the maximum distance that a particle can take in an iteration. Speed update of a particle can be differentiated for each dimension $j \in 1 \ldots n$ (n are based on number of optimized parameters), so $v_{i,j}$ represent the dimension to j from speed vectors that associated with particle to i. So the equation for speed (Bai, 2010):

$$v_i = \dot{\omega} v_i + c_1 r_1 (pbest - x_i) + c_2 r_2 (gbest - x_i) \quad (5)$$

Explanation :
vi   : curent particle velocity
ω   : weight of inertia
c1, c2 : acceleration coefficient
r1, r2 : uniform random number between 0 and 1
gbest : current global best position
xi   : curret particle position

5. Evaluation, aims to create accuracy and kappa valuess of the model that has been succesfully created in the previous step. To know the level of accuracy of the resulting models, using confusion matrix based on X-Fold Cross Validation method.
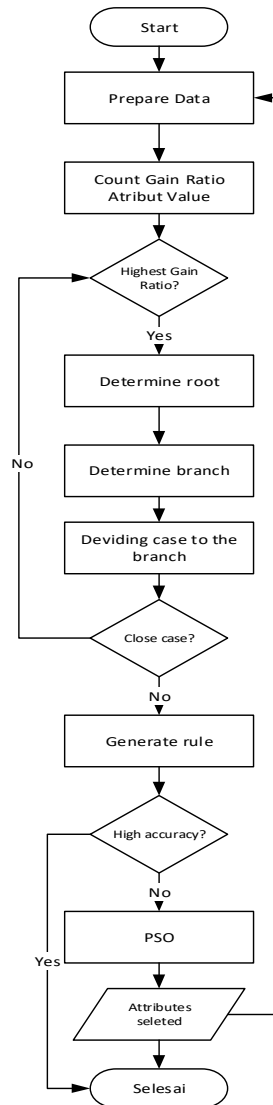6. Deployment, where the result of all previous stages are use in real.

Fig. 2. C4.5 Algorithm Flowchart.

The flowchart (Fig.1) is combines the process of the C4.5 algorithm with the PSO. If the result of algorithmc4.5 get a less good accuracy, then proceed with the PSO process. After that, the result from PSO process be data input in C4.5 algorithm process so as to produce pattern shown as decision tree.

3. Result And Discussion

3.1 Business Understanding

The purpose of this research is to know accuracy of C4.5 algorihtm with PSO for feature selection, by using the sample data of performance result with soft competencies of teacher for identify the attribute that have an important role in modelling. In expected by using PSO, can improve the accuracy of form generated by C4.5 algorithm.

Mapping the soft competencies of teacher by psycho test using measuring instrument in psychology, while performance appraisal is done based on the aspect of hard competencies. It requires the knowledge of linkages between soft and hard competencies to specific profession with its various pecularity. Based on processing result by using C4.5 algorithm then the resulting format can used to know the relation between soft competencies and teacher performance.

3.2 Data Understanding

The data used in this research is to know soft competences feature for teacher performance of C4.5 algorithm accuracy in classifying soft competencies data and teacher performance. The data abtained from teacher result assessment and teacher perfomance appraisal.

Table 1. Teacher Soft competencies.

No	Soft Competencies	Code
1	Conceptual Thinking	CT
2	Analitycal Thinking	AT
3	Achivement Orientation	AO
4	Concern for Order	CfO
5	Planing and Organizing	PnO
6	Interpersonal Understanding	IU
7	Customer Service Orientation	CSO
8	Flexibility	F
9	Communication	C
10	Organizational Commitment	OC

From table 1 we can explained that in research data there is data of teacher performance appraisal result that can become research label and 10 attribute of teacher soft compentencies with each assessment has a number from 1 up to 6. Each of teacher get the teacher performance rating between 0 up to 100, as presented in table 3 below:

Table 2. Criteria of teacher performance rating:

Criteria	Initial	Scala
Very Good	A	91 – 100
Good	B	76 – 90
Fair	C	61 – 75
Poor	D	0 – 60

The provisions from HRD section, that the scoring scale is as included in table 2. To get a grade A, the teacher should get a total score above 90. Assessment data result and

performance appraisal that was sample in this research is data of 133 teacher taken from kindergarten, elementary school, junior high school, and senior high school.

3.3 Data Preparation

In the data preparation stages, to get data for use in implementation research , do the selection of data from available data based on completeness of data on the attribute used, there are 163 data. Through on the process of cleaning the data, there are some incomplete data, so resulting 133 data that can be used as research data. Data processed by using software rapid miner.

Based on table of calculation gain ratio of C4.5 algorihtm, can be shown the scale of gain ratio as follows:

Table 3. Gain Ratio for each attributes.

No	Atribut	Gain Ratio
1	Achivement Orientation (AO)	0.22
2	Planing and Organizing (PnO)	0.21
3	Concern for Order (CfO)	0.20
4	Customer Service Orientation (CSO)	0.20
5	Analitycal Thinking (AT)	0.19
6	Conceptual Thinking (CT)	0.18
7	Organizational Commitment (OC)	0.16
8	Communication (C)	0.15
9	Flexibility (F)	0.13
10	Interpersonal Understanding (IU)	0.09

From table 3 about the gain ration,it can be explained that highest gain ratio value is in AO attribute. OA attribute becomes root in making decision tree, then followed by another attributes according the value of each gain ratio.

3.4 Modelling

Of the 10 attributes in teachers' soft competencies and teacher performance, the C4.5 algorithm can be shown the decision tree as follows (Figure 2):
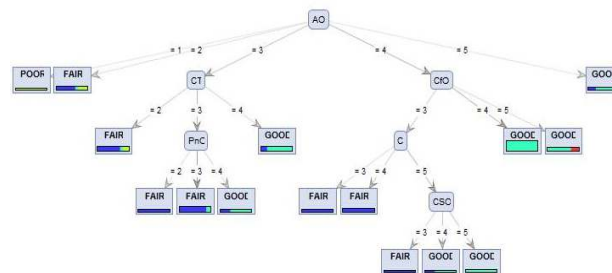

Fig. 3. Result models of data classification using C4.5 algorithm.

By using C4.5 algorithm, we get the patterns as shown on figure 2. In the resulting models can explain that the attribute has highest gain ratio in root position is AO (Achievement Orientation). AA become the attribute that has high influence in making the resulting classification pettern to determine the teacher performance. If the AO score is 1 so the rating result is poor, if the AO score is 2 so the rating result is fair, if the AO score is 3, the

7

prediction of teacher performance appraisal result depends on the value of CT and Pno, if the OA score is 4, the performance appraisal will see the value of C and CSO. Classification patterns that generated from C4.5 algorithm, can be known the accuracy levels by using X-Cross Validation technique. The generated accuracy is 70,83%.

Table 4. Confusion matrix classification by C4.5 Algorithm.

	true FAIR	true GOOD	true POOR	true VERY GOOD	class precision
pred. FAIR	30	11	6	0	63.83%
pred. GOOD	11	50	1	2	78.12%
pred. POOR	2	0	0	0	0.00%
pred. VERY GOOD	0	0	0	0	0.00%
class recall	69.77%	81.97%	0.00%	0.00%	

Accuracy: 70.83% +/- 16.14% (mikro: 70.80%)

The accuracy declares the true level of resulting models. With class precision of fair prediction is 63.83%, good prediction is 78.12%, very good and poor prediction is 0%. From 64 good predicted data, 48 data is true (good), 16 data is false (bad). While the class recal of fair prediction is 69.77%, good prediction is 81.79%, very good and poor prediction score is 0%.

Table 5. Kappa table Classification by C4.5 Algorithm.

	true FAIR	true GOOD	true POOR	true VERY GOOD	oclass precision
pred. FAIR	30	11	6	0	63.83%
pred. GOOD	11	50	1	2	78.12%
pred. POOR	2	0	0	0	0.00%
pred VERY GOOD	0	0	0	0	0.00%
class recall	69.77%	81.97%	0.00%	0.00%	

Kappa: 0.449 +/- 0.287 (mikro: 0.454)

In confusion matrix table (table 5) the kappa values for C4.5 algorithm classification result is 0.449. The value of kapaa is on sufficient category.

In experimental of data classifiaction using C4.5, from the table 4 and 5, it can be shown that C4.5 algorithm has accuracy 70.83% and kappa values 0.449, so result from this data experiment is model is good level.

The implemantation of PSO feature selection use to get the value of attribute weight used in dataset. The implementation process of PSO for C4.5 algorithm can shown as the pictures 3 as below:
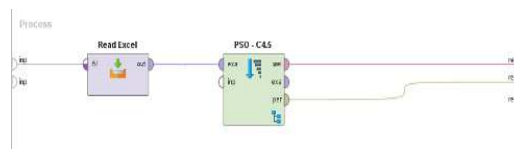

Fig. 4. PSO process design for C4.5 algorithm.

8

In process of atribute selection using PSO, used one occurent design for C4.5 algorithm, by using the Read Excel operator. Multiplay to use of one family, PSO are used to try the results of accuracy PSO application in algorithm C4.5. In the cross validation operator , used C4.5 algorithm as the algorithm making model, Apply model and Performance.

From the process of PSO conducted on the teacher performance appraisal and teacher assessment, in C4.5 algorithm used prepunning and postpruning technique, obtained atribute weighting as follows (table 6). Pruning technique is operated to get the sample patterns, by removing the attribute that produced the branch with little effect, based on PSO calculation.

Table 6. The result of C4.5 algorithm feature selection by using PSO.

Atribute	Weight
AO	1.0
F	1.0
CT	0.4821792213995124
CfO	0.26197811520778463
OC	0.12492439951584233
C	0.11462660697205894
AT	0.0
PnO	0.0
IU	0.0
CSO	0.0

Based on table 6 can be explained that the attributes in the dataset used to create the classification patterns which has a variety of weight value. There is an attribute with a recommendation value to use as an attribute in creating the classifications patterns is for the OA, F, CT, CfO, OC and C attributes. While the other attribute has a small weight value and not recommended to use as an attribute in classification process of model formation, thera are AT, Pno, IU, and CSO attributes. In process of features selection using PSO to used dataset in create the classification patterns by using C4.5 algorithm, which have 79,70% accuracy.

Table7. Confusion Matrix of accuracy value in attribute selection using PSO for C4.5 algorithm.

	true FAIR	true GOOD	true POOR	true GOOD	VERY	class precision
pred. FAIR	37	8	7	0		71.15%
pred. GOOD	6	53	0	2		86.89%
pred. POOR	0	0	0	0		0.00%
pred. VERY GOOD	0	0	0	0		0.00%
class recall	86.05%	86.89%	0.00%	0.00%		

Accuracy: 79.70% +/- 7.93% (mikro: 79.65%)

With class precission of fair prediction is 71.15%, good prediction is 85.89%, very good and poor predictions score is 0%. From 61 good predicted data, 53 data is true (good), 8 data is false(bad). While the class recal of fair prediction is 86.05%, good prediction is 86.89%, very good and poor prediction score is 0%.

Table 8. Confesion Matrixo of kappa values in attribute selection using PSO and C4.5 algorithm

	true FAIR	true GOOD	true POOR	true GOOD	VERY	class precision
pred. FAIR	37	8	7	0		71.15%
pred. GOOD	6	53	0	2		86.89%
pred. POOR	0	0	0	0		0.00%
pred. VERY GOOD	0	0	0	0		0.00%
class recall	86.05%	86.89%	0.00%	0.00%		

Kappa: 0.528 +/- 0.257 (mikro: 0.618)

Based on Confusion Matrix table (table 8), the kappa valuess for attribute selection using PSO and C4.5 algorithm is 0.528. The kappa valuess is on sufficient category.

In experiments by added feature selection C4.5 algorithm, C4.5 uses PSO accuracy is 79.70%. To get the patterns of C4.5 algorithm with PSO, the result of attribute selection is applied to C4.5 algorithm by producing this following patterns:
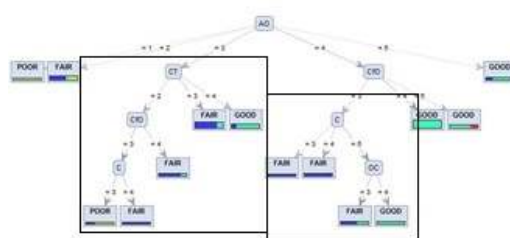

Fig. 5. The patterns of data classification result C4.5 using PSO attribute selection.

The pattern in the box on fig.4 is different result from C4.5 algorithm with PSO. On the patterns that produced by C4.5 algorithm using PSO attribute selection, the highest influence attribute is AO (Achievement Orientation). AO can be the highest influence in generating the classification result to determine the teacher performance. If the AO value is 1, the rating result is poor, if the AO value is 2, the rating result is fair, if the AO value is 3, the prediction of teacher performance appraisal result depends on the value of CT, Pno and C, if the OA score is 4, the performance appraisal will see the value of C and CSO. If the OA value is 5, prediction of teacher performance appraisal is good.

3.5 Evaluation

The fifth step of the research method is evaluation. From the experiments process by using C4.5 algorithm applied in PSO to get the recommended attribute based on weight result from PSO process.

Based on the experiment of accuracy and kappa valuess, data classification using C4.5 algorithm with PSO respectsively in PSO feature selection, C4.5 algorithm with PSO produced the higher accuracy value that is 79.70% and kappa values 0.528 (sufficient category).

Table9: Comparison of accuracy and kappa valuess C4.5 algorithm and rule induction.

Algorithm	Accuracy	Kappa

C4.5	70.83%	0.449
C4.5 with PSO	79.70%	0.528

In experimental of data classifiaction using C4.5 and C4.5 with PSO, from the table 9 it can be shown that C4.5 algorithm with PSO has the highest of accuracy.To get the better accuracy from the experiments performed C4.5 with PSO is 79.70% using particle swarm optimization (PSO) to select the attribute used.

3.6 Deployment

At the deployment step, communicate with user who will use the results of research. Based on result of research, in the deployment will use tha result of C4.5 algorithm with PSO. To discuss to the user related to the implementation of the research result.

4. Conclusion

From the result of research, obtained that the processing data of performance and soft competencies of teacher by using C4.5 algorithm with PSO produce the accuracy on 79.70%. The kappa valuess for C4.5 algorithm with PSO is 0.528. From the result of accuracy and kappa, can be concluded that in classification data of teacher for performance aspect and soft competencies. The dominant factor that influence teacher performance base on soft competencies is achievement orientation, then conceptual thinking and concenr for order, communication and organizational commitment.

For further research, research can be done using different classification algorithms such as C5, rule induction, naive bayes, K-NN and others. Other research can be continued with the design of the resulting pattern implementation.

Refferences

[1]   Abadi, L. P. (2016). Kajian Penerapan Algoritma C4 . 5 , Naive Bayes , Dan Neural Network Dalam Pemilihan Dosen Teladan : Studi Kasus Universitas Indraprasta, 9(3), 237–246.

[2]   Aggarwal, Charru, 2015, Data Mining: The Textbook, Ibm T.J. Watson Research Center Yorktown Heights New York Usa

[3]   Alaydroes, Fahmy, dkk. 2010, Standar Mutu Jaringan Sekolah Islam Terpadu Indonesia, Jakarta, JSIT

[4]   Asanbe, M. O., & William, W. F. (2016). Teacher's Performance Evaluation in Higher Educational Institution using Data Mining Technique, 10 (7), 10–15.

[5]   Bai, Q, 2010. Anaysis of Particle Swarm Optimization Algorithm. Computer and Information Science-CCSE, 180-184

[6]   Galway, N. U. I. (2011). An Exploration of Gradually Increasing Directed Neighbourhoods for Particle Swarm Optimisation, 0–19.

[7]   Garikapati, P. T., Penki, N. K., & Gogineni, S. (2017). Improvised Gene Selection Using Particle Swarm Optimization With Decision Tree As Classifier, (9), 80–86.

[8]   Gupta, S., Kumar, D., & Sharma, A. (2011). P ERFORMANCE A NALYSIS OF V ARIOUS D ATA M INING C LASSIFICATION T ECHNIQUES ON, 3(4), 155–169.

[9]   Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003). Support Vector Machine.

[10]   Larose D.t, 2005. Discovering Knowledge in Data. Canada: Wiley-Interscience

[11]   Rizal, Y., Hubeis, M., Mangkuprawira, S., Keuangan, D., Manajemen, D., &

Pascasarjana, S. (2013). Pengaruh Faktor Kompetensi Terhadap Kinerja Individu di Perusahaan Agroindustri Go Public, 8(1), 1–8.

[12] Samad, A., Basari, H., Hussin, B., Pramudya, I. G., & Zeniarja, J. (2013). Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. Procedia Engineering, 53, 453–462. https://doi.org/10.1016/j.proeng.2013.02.059

[13] Sanjay, S. S., & Keshav, B. B. (n.d.). Teacher' s Performance Analyzer, 2 (D).

[14] Utaminingsih, Sri. 2015, Pengembangan Kompetensi Guru Dalam Manajemen Pembelajaran Kurikulum 2013, FKIP Universitas Muria, Kudus.

[15] Tsai, M., Chen, K., Su, C., & Lin, H. (2012). An Application of PSO Algorithm and Decision Tree for Medical Problem, 124–126.

[16] Wu, Xindong And Vipin Kumar, 2009, The Top Ten Algorithm In Data Mining, Crc Press, Taylor & Francis Group, Llc