Kontigensi: Scientific Journal of Management
Vol 10 , No. 2, Desember 2022, pp. 431 - 438
ISSN 2088-4877

Analysis of Self-Organizing Maps and Fuzzy C-Means methods in Clustering Teacher Data for Nominations of Candidates for Education Unit Supervisors

Abd. Rahman
UPT TIPD IAIN Palopo, Indonesia
E-mail: abdrahman@iainpalopo.ac.id

ABSTRACT

The quality of education is very dependent on the management of education management one of the important factors in the management of education is monitoring and evaluation. Therefore, continuous Education Supervision supported by the appointment of competent supervisors will have implications for the quality of education. Supervision of education management is carried out by the principal or teacher appointed to carry out managerial and academic supervision in certain educational units. The problem that often occurs is that grouping teacher data to be selected as supervisor candidates are still conventional. Therefore, a teacher data grouping model is needed to obtain useful information in planning strategic steps and policy regulations for determining prospective supervisors for academic units. By utilizing Information and Communication Technology, especially in the field of Data Mining, this Teacher Data Grouping research uses the Fuzzy C-Means and Self Organizing Maps method, and the grouping results are analyzed by measuring the spread of data on each grouping formed by using cluster variance. The output of the Teacher Data grouping process using the Fuzzy C-Means and Self Organizing Maps methods can bring up a group of Teacher Data nominations that are competent to be selected as candidates for supervisors for certain educational units. The research results were obtained by forming several groupings in Fuzzy C-Means by providing an error accuracy value of 0.1 and Self Organizing Maps, which is set by the learning rate and learning rate; the results obtained are grouping with 3 clusters by providing a learning rate of 0.8 and a learning rate of 0.7 in the Self Organizing Maps method has a Variant value that is ideal compared to grouping on Fuzzy C-Means and rather than the same method by forming different groups.

Keywords: Cluster Variance, Clustering, Fuzzy C-Means, Self-Organizing Maps, Teacher Data, Supervisor.

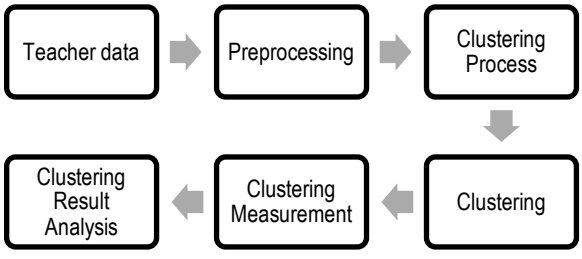
INTRODUCTION

Education is an important factor in achieving the level of welfare of a nation. Quality education will produce quality and independent community output to improve the quality of education; in addition to management factors, monitoring and evaluation are needed by conducting supervision at every level of the education unit. In the context of education implementation, the concept of supervision occupies a very strategic position because no matter how good an educational program planning is, if an adequate supervision process does not accompany it, then all previously launched programs will not measure the level of success, even very possible. Some deviations occur in it that are difficult to detect.[9]. It is necessary to appoint a supervisor in each education unit for supervision. Education unit supervision is a school principal and teacher given the complete task, responsibility, and authority to conduct educational supervision by conducting assessments and guidance from the technical aspect of education and administration in certain educational units [9]. Continuous supervision of education management and the appointment of competent supervisors will have implications for the quality of education. The problem currently happening is that the process of grouping teacher data to be selected as supervisor candidates is still carried out conventionally, unable to explore teacher data

quickly and accurately, and non-objective assessment of candidates according to the competence of each prospective supervisor. So that a grouping of teacher data is needed to get useful information in planning strategies and policy steps in making decisions for the acceptance of candidates for supervisory education units; by utilizing information and communication technology, especially in the field of Data Mining, this study will use the Fuzzy C-Means (FCM) and Self Organizing Maps (SOM) methods; both methods are used to obtain an ideal grouping model for the characteristics of teacher data so that the results of the grouping of teacher data can be obtained that can be recommended to be appointed as candidates for supervisors of academic units by the criteria for assessing the criteria of each parameter of teacher data accurately to become candidates for supervisors of academic units at certain levels.

METHOD

The method that will be used in this research is by utilizing information and communications technology, especially in the data mining sector, for data clustering, using the Fuzzy C-Means Method and Self Organizing Maps; this is by the various characteristics of teacher data that need to be clustered so that it can bring up prospective supervisors who have the competence and are eligible to be nominated as supervisors, these two methods are classified as unsupervised learning, meaning unsupervised learning, in the Artificial Neural Network concept, the output is not determined by the target to be achieved. Therefore, the work system in this research is as follows:



Picture 1. Research Block Diagram

The data used in this research is from teachers and principals from the North Aceh district in the form of quantitative data. Researchers get data by collecting data from various types of data sources, such as teacher competency test data, teacher certification data, and functional position allowance data so that sufficient data is obtained in this study. The teacher data attributes used are the teacher/principal status, education level, educational qualifications, years of service, age, rank, and certification. In addition, it is by the parameters of the assessment rules for prospective education unit supervisors. The amount of data used is 315 people.

Before the teacher data enters the clustering process stage, it is first processed at the preprocessing data stage; this is done for data validation against duplication and cleaning data records that are not used, and then normalization is carried out in the form of data transformation to be used in the grouping process by adjusting the transformation input data between the range 0 to 1, using the following formula:

$$\mathrm{x}^{n} = \frac{x - \min(x)}{\max(x) - \min\,(x)}$$

Specifically for parameters, status, educational qualifications, education level, rank, and certification, this is done by converting the parameter data into a range of 0 and 1; for example, the status parameter is converted as follows:

Table 1

Status	Conversion	Input Value
Principal of Vocational High School	20	1
Principal of Senior High School	19	0,95
Teacher at Vocational High School	18	0,90
Teacher at Senior High School	17	0,85

Principal of Junior High School	16	0,80
Teacher at Junior High School	15	0,75
Principal of Elementary School	14	0,70
Teacher at Elementary School	13	0,65
Principal of Kindergarten	12	0,60
Teacher at Kindergarten	11	0,55

The real data collected is then carried out with data preprocessing actions. This activity is intended to clean data from data that is not needed. After the data preprocessing stage is complete, proceed to the clustering stage using the Fuzzy C-Means method and Self Organizing Maps:

1. Fuzzy C-Means (FCM)

Clustering with FCM Method is one of the methods in Artificial Neural Networks, which is unsupervised learning based on fuzzy logic theory. The basic concept of FCM is to determine the center point of the cluster to be marked as the average location point for each cluster. For example, the following is the algorithm of FCM:

a. Determine the data in the form of matrix nxm, number of clusters (c≥2), weighting power (w>1), maximum iteration, the objective function (P0=0), accuracy (ε=very small positive value), initial iteration (t=1).
b. Generate initial partition matrix (μ) randomly
c. Calculate the cluster center of each cluster

$$V_{kj} = \frac{\sum_{j=1}^{n}((\mu_{ik})^{w} \times \times_{ij})}{\sum_{j=1}^{n}(\mu_{ik})^{w}}$$

d. Calculate the objective function in the –t iteration

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}([\sum_{j=1}^{m}(x_{ij} - v_{kj})^2](\mu_{ik})^{w})$$

e. Update membership of μ degree

$$\mu_{ik} = \frac{[\sum_{j=1}^{m}(x_{ij} - v_{kj})^2]^{\frac{-1}{\omega-1}}}{\sum_{k=1}^{c}([\sum_{j=1}^{m}(x_{ij} - v_{kj})^2]^{\frac{-1}{\omega-1}}}$$

with i=1..2..n ; k=1..2..c

f. Check the stop condition
   1) If : (P1 – Pt) < ε or (t < MaxIteration) then stop,
   2) b. If not: t = t +1, repeat step 3.

In clustering using the FCM method, the initial partition matrix value (μ) will be used randomly, and the error accuracy value is 01; this value is given to each group to be formed, namely 2, 3, and 4 groups, and the difference in the results of the data distribution will be seen.

2. Self-Organizing Maps (SOM)

Like the FCM method, the SOM method is also one of the methods in the Artificial Neural Network that is unsupervised learning, where no target data is specified. In SOM, basic knowledge is given in the form of data parameters and specified weights; only the winning neurons are updated with their weights. The algorithm for the SOM method is as follows

a. Initialization:
   1) Determine the weight of W_{ij} randomly,
   2) Determine the learning rate,
   3) Determine the shape and radius (R).
b. If the termination condition is false, perform steps 3 to 8,
c. For each input vector X, perform steps 4 to 6,
d. Calculate the Euclidian distance $D_{(j)}$, for each j (j=1..2..m) with the weight value W_j and the input data X_i (1=1..,2..,3), with the equation:

$$D_j = ||W_j - X_i||^2 = \sum_i (W_j - X_i)^2$$

e. Determine the index j such that D_j is minimum
f. Fix the value of the W_{ij} Weight for each unit j around J by using the equation:

$$Wij\_new = W_{ij\_old} + \eta\,(X_i - W_{ij\_old})$$

g. Modify the learning rate when the iteration is complete:

$\eta_{(t+1)}=0{,}5_{(t)}$

h. Test the stopping condition.

To get the maximum data distribution results, the learning rate will be determined in the SOM clustering method, and each data cluster formed will be distinguished by the learning rate and learning rate values.

3. Clustering Evaluation Analysis

The resulting clustering of teacher data will be evaluated for clustering data by analyzing the measurement of the spread of data that occurs using cluster variance, which is a model for measuring the distribution of data that occurs. In a cluster variance, a cluster reaches the ideal if the membership value of a group has a high similarity and is different from other groups. The magnitude of the spread data value that occurs in a group can be formulated as follows:

$$V_c^2 = \frac{1}{n_c - 1}\sum_{i=1}^{n_c}\left(d_i - \overline{d_i}\right)^2$$

Where:
V_c^2 = variance in c cluster
c = 1..k, where k=number of clusters,
d_i = i data in a cluster,
di = the average of data in one cluster,

In cluster variance, there are two models of clustering evaluation measurement from an internal perspective, namely Variance within Cluster (Vw), which is used to measure the level of data spread within a group that is formed; the ideal group has a minimum Vw value, which can be formulated as follows

$$V_w = \frac{1}{N-k}\sum_{i=1}^{k}(n_i - 1).V_i^2$$

Where:
N = sum of all data
n_i = the amount of data in a group i,
V_i^2 = variant on cluster i

Furthermore, from an external perspective, the measurement of cluster variance is Variance Between Clusters, which is a measurement of the spread of data between clusters that are formed; a high value of *Vb* indicates an ideal cluster, which can be formulated as follows:

$$V_b = \frac{1}{c-1}\sum_{i=1}^{c} n_i\left(\overline{d_i} - \overline{d}\right)^2$$

Where:
c = number of clusters
n_i = amount of data in cluster i
d = average value d_i

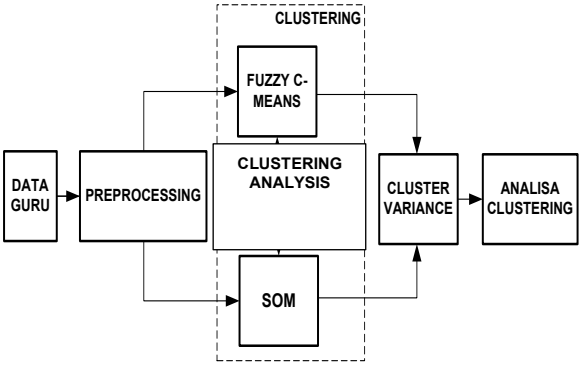
The overall variance of all groups formed can be calculated by calculating the Variance Within Cluster (*Vw*) and Variance Between Clusters (*Vb*); the smaller the value of a Variant (*V*) indicates an ideal group to be able to calculate the variance of the entire group formed is with the following formula:

$$V = \frac{V_w}{V_b}$$

Based on the analysis of the clustering of teacher data using the FCM and SOM methods, the ideal clustering results will be taken in terms of measuring cluster variance to become a recommendation for the selection of candidates for supervisors of academic units.

RESULTS AND DISCUSSION

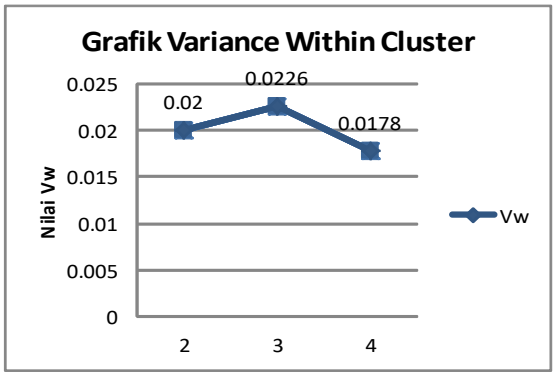
The following is the clustering flow that is carried out:



Picture 1. Research Workflow

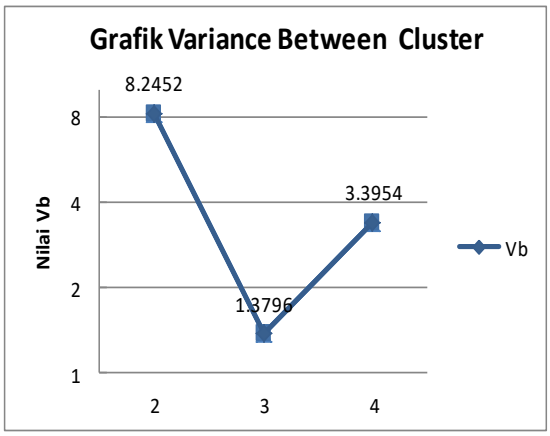
1. Fuzzy C-Means (FCM)

The results of clusters with a minimum variance value with cluster formation > 2 in the FCM will be used as the center of the cluster as the initial weight for the SOM method for further clustering with the SOM method. Giving weight to the SOM method in the form of a cluster center in the FCM method is carried out to measure how much value the data spread by adjusting the learning rate, which will be compared with clustering using the SOM method, where the weight value is determined randomly. In data clustering, the teacher uses FCM to form 2, 3, and 4 clusters by setting the parameters, namely the error accuracy value = 0.1, maximum iteration = 100, objective function = 0, weighting power = 2, and for each data membership value or μ the matrix is a random value. The value information is obtained for the data distribution within clusters and between clusters, and the overall variance value is as follows:



Picture 2. Value of Variance Within Cluster FCM

The results of measuring the spread of data by forming 2, 3, and 4 clusters on FCM obtained a grouping with 4 clusters being the most ideal, but the variance between clusters must be known, while the *Vb* value is as follows:



Picture 3. *Vb* value on FCM

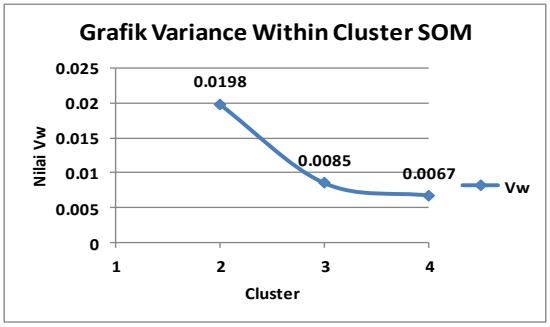
The best overall variance value is grouping using the FCM method on 2 clusters with a variance value of 0.00243. Clustering with 2 clusters in the FCM for selecting prospective supervisors for education units has not been able to represent prospective supervisors who meet the nominations for the main, medium, and low candidates because the best grouping is in the cluster formed with 2 clusters. So the 4 clusters will be taken with a variance value of 0.00524, with the value of the cluster center as follows:

Table 2

0.77	0.53	0.95	0.79	0.43	0.85	0.89
0.75	0.53	0.96	0.81	0.50	0.86	0.92
0.75	0.53	0.96	0.81	0.50	0.86	0.91
0.77	0.54	0.95	0.78	0.44	0.85	0.88

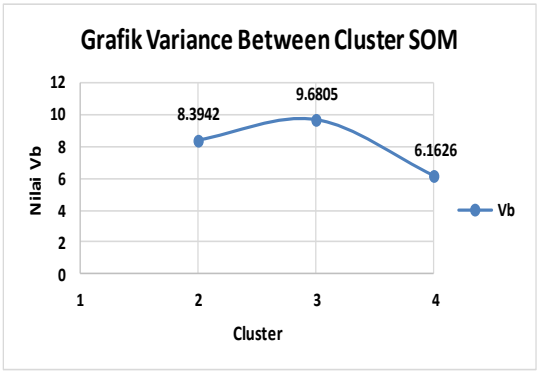
2. Self-Organizing Maps (SOM)

In the clustering results obtained using the SOM method, the initial weight values are given randomly, having different variance values, where for clustering 2 clusters with a learning rate of 0.6 and a learning rate of 0.5, for 3 clusters a learning rate of 0.8 and a learning rate of 0.7 are given, and for 4 clusters given a learning rate of 0.9 and a learning rate of 0.8. The following are the results of the cluster variance in the form of graphs of *Vw, Vb,* and the overall variance, namely:



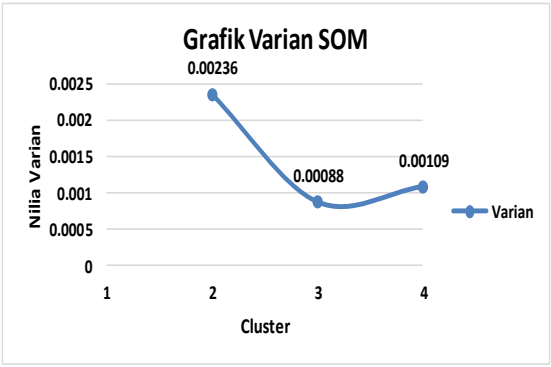
Picture 4. *Vw* value on SOM

The Variance Within Cluster (*Vw*) value in SOM has the best value in clustering 4, with a *Vw* value of 0.0067. Furthermore, the results for Variance Between Clusters (*Vb*) for clustering 4 clusters do not have the best value but exist in clustering with 3 clusters, and the following is the value of Variance Between Clusters (*Vb*) on SOM:



Picture 5. *Vb* Value on SOM

Based on the Variance Between Clusters (Vb) value, the maximum value is in clustering 3 clusters with a value of 9.6805. Therefore, the overall variance value in SOM clustering is as follows:



Picture 6. Variance Value on SOM

In clustering using the SOM method, information was obtained that clustering by forming 3 clusters with a learning rate of 0.8 and a learning rate of 0.7 has an ideal variance result compared to clustering for 2 and 4 clusters with SOM with a variance value of 0.00088 which is more than clustering in the SOM method. FCM.

3. FCM + SOM

The value of the cluster center will be used as the initial weight for grouping with the combination of FCM and SOM methods; by providing a learning rate of 0.9 and a learning rate of 0.8, the Cluster Variance value is obtained as follows:

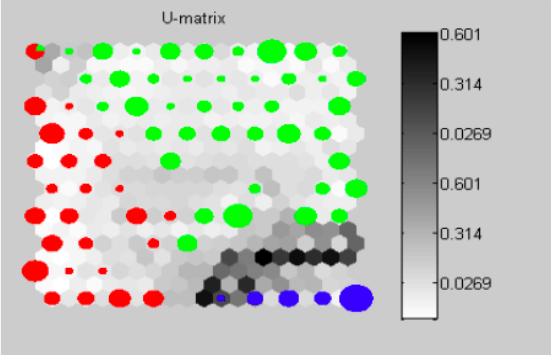
Table 3

Evaluation	Value
Variance Within Cluster	0.0068
Variance Between Cluster	6.0741
Overall Variance	0.00112

It turns out that the variance value in FCM+Kohonen SOM has more ideal results than grouping in FCM by forming 2, 3, and 4 clusters.

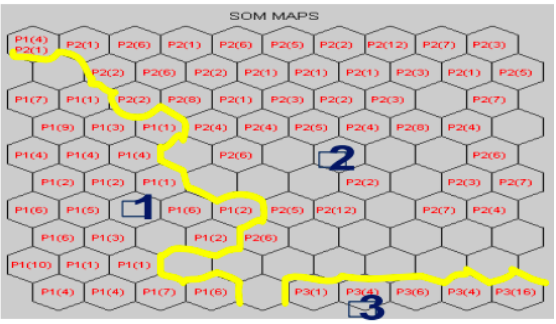
4. Visualization

The results of clustering self-organizing maps can be represented as a unified distance matrix or u-matrix. Grouping is visualized on clusters with the best variance value; this is done to see the nominations of each input data to be selected in the main, medium, and low nominations. Furthermore, here are the results of the u-matrix in grouping by forming 3 clusters as follows:



Picture 7. U-matrix Visualization

The results of color degradation, which are mapped with the u-matrix, show the position of each input data that enters the grouping formed by 3 clusters. Furthermore, the entire data entered in each grouping formed as follows:



Picture 8. U-matrix Visualization and Labeling

The results of the u-matrix visualization can be broken down into the input data included in the grouping as follows:

Cluster	Teacher Data		
	Main Nomination (P1)	Medium Nomination (P2)	Low Nomination (P3)
1	105	4	-
2	-	175	-
3	-	-	31

There are 105 main nominations (cluster 1), 179 moderate nominations (cluster 2), and 31 low nominations (cluster 3). Based on the results of research conducted, the minimum variance value is in clustering by forming 3 clusters of the SOM method with random initial weights, by providing a learning rate of 0.8 and a learning rate of 0.7, with the results of the variance within cluster (Vw) of 0.0085 and the variance between clusters (Vb) is 9.6805, and the overall variance value is 0.00088, which is the minimum variance value. Furthermore, the process of distributing data from 315 teacher data, obtained information for cluster 1 as many as 105 teacher data, cluster 2 as much as 179 data, and 31 data entered cluster 3. The results of this clustering can represent from the main, medium, and low nominations to be used as recommendations for selecting candidates for supervisors of education units.

CONCLUSION

Based on the results of the research conducted, several things can be concluded, namely as follows:

1. The clustering method with FCM provides an error accuracy of 0.1, the value of U-matrix data membership randomly by forming 2, 3 and 4 clusters have an excellent data distribution value in clustering 2 clusters with a value of 0.02, and the spread value between clusters the best in the grouping of 2 clusters of 8.2452. the best overall variant is also in grouping 2 clusters of 0.00243.
2. The clustering method using SOM by forming 2,3 and 4 clusters and assigning learning rates and learning rates to each cluster that is formed, has a cluster variance value that is more ideal than FCM and clustering the combination of FCM with SOM, this is based on the overall variance value in grouping with 3 clusters, which is 0.00088.
3. The value of learning rate and learning rate on the SOM method determines the desired data distribution results according to the formed cluster.

In the development of future research, it is necessary to pay attention to the error accuracy value determined in the FCM, and the SOM method requires a special determination to pay attention to the initial weight matrix value, it is also necessary to evaluate the measurement of grouping results with other methods.

REFERENCES

Ana Ratna Wati, D., & Agusti Rochman, Y. (2008). *Model Penjadwalan Matakuliah Secara Otomatis Berbasis Algoritma Particle Swarm Optimization (PSO)*. UII Yogyakarta.

Aulia, I., Nababan, E. B., & Muchtar, M. A. (2012). Penerapan Harmony Search Algorithm dalam Permasalahan Penjadwalan Flow Shop. *Jurnal Dunia Teknologi Informasi*, *1*, 1–7.

Desiani, A., & Arhami, M. (2006). *Konsep Kecerdasan Buatan*. Andi Offset.

Geem, Z. W., Kim, J. H., & Loganathan, G. V. (2001). A New Heuristic Optimization Algorithm: Harmony Search. *Simulation*, *76*(2), 60–68. https://doi.org/10.1177/003754970107600201

Mahdavi, M., Fesanghary, A., & Damangir, E. (2007). An. Improved Harmony Search Algorithm for Solving Optimization Problems. *Applied Mathematics and Computation*, *188*, 1567–1579.

Priyo, S. S., Helmie, A. W., Indra, W., Nurdin, B., & K, T. W. (2016). *Chapter 7 Backpropagation dan Aplikasinya* (pp. 135–146).

Rachmawati, H. (2012). *Analisis Penyelesaian Masalah Penjadwalan Kuliah Menggunakan Pewarnaan Graph oleh Algoritma Koloni Lebah dengan Parameter Soft Contsraints Prioritas Dosen*. ITS Surabaya.

Santosa, B., & Willy, P. (2011). *Metoda Heuristik, Konsep dan Implementasi.* Guna Widya.

Setemen, K. (2008). *Kombinasi Algoritma Genetika dan Tabu Search dalalm Penyusunan Jadwal Kuliah*. ITS Surabaya.

Sutiono, N., Imanuel Malik, N., & Rojali. (2012). *Automatisasi Timetabling Asisten Pengajar pada Software Laboratory Center menggunakan Harmony Search*. Binus University.

Suyanto. (2007). *Artificial Intelligence; Searching, Reasoning, Planning and Learning*. Informatika.

Teddy. (2009). *Penyelesaian Penjadwalan Kuliah Sebagai Constraint Satisfaction Problem dengan Genetic Algorithm*. Universitas Indonesia.