Journal of Engineering and Applied Sciences 14 (11): 3630-3637, 2019
ISSN: 1816-949X
© Medwell Journals, 2019

Zone Based School Enrollment System by using Modified k-Means Clustering Method

Purba Daru Kusuma
Department of Computer System School of Electrical Engineering,
Telkom University Buah Batu Street, Bandung, 40257 West Java, Indonesia

Abstract: The main goal of zone based school enrollment method is reducing the student's home to school distance. So, among its variants, student will be allocated to the school that its location is near the student's home location. Meanwhile, one biggest problem in this method is the number of out of zone students which means of the student that is accepted in school that its location is outside his own zone. This situation occurs, especially when the zoning is determined statically. In this research, we propose dynamic zone based school enrollment model to meet the variations in school location distribution, student location distribution and school capacity. This model is developed based on k-means clustering method with several modifications. This proposed model then is implemented into zone based school enrollment simulation to measure its performance. In this simulation, the performance of this proposed model is also compared with the existing static one. In this simulation, adjusted parameters are school location distribution, student location distribution, school capacity, number of schools and number of students. Meanwhile, the observed parameters are average student's home to school distance and the success ratio. Based on the test result, there are several findings. The proposed model produces higher success ratio rather than the existing static method does. In the other side, the existing static method produces lower average student's home to school distance rather than the proposed model does. The number of schools and the average school capacity has positive correlation with the success ratio. The number of students has negative correlation with the success ratio. These adjusted variables do not affect the average student's home to school distance.

Key words: Student enrollment, zone based model, k-means clustering, variants, static method, adjusted variables

INTRODUCTION

Zone based school enrollment method is one popular enrollment method that has been adopted in many countries such as New Zealand (Devonport, 2017; Beaven, 2003; Boser, 2013) America (Kalata and McEwen, 2015; Hill, 1990; Hastings *et al.*, 2007; Fossey, 1994) and Australia (Banks, 2005). Even this method has been applied in many places, debates about this method still occurs. In his research, Fossey (1994) said that families generally enrolled their children in the schools with higher indicators of student performance and socioeconomic status. In other research, parents with higher income and have children with high academic ability preferred schools with higher academic score (Hastings *et al.*, 2005).

Now a days, this method is also adopted in Indonesia, especially in this current administration (Dewi and Elyda, 2013). Based on the regulation that is announced by the ministry of education and culture, public schools must implement the zone based method for most of students that are accepted (Dewi and Elyda, 2013). The main goal of the government of Indonesia to implement the zone based system is creating academic performance equality among schools (Dewi and Elyda, 2013). The reason is when the school selects new students based on their academic performance for example is based on the national exam score, smart students will tend to enter high quality public schools. So, higher quality public schools get smarter students. In the other side, lower quality public schools are entered by less smart students. Based on this wide input gap, the the academic performance gap among schools is still wide. So, by distributing the smart students among all schools, the government hopes that the academic quality gap among schools can be reduced or eliminated.

The other goal of this method is reducing student home to school distance. It is common that some students or parents tolerate farther home to school distance to get higher quality school. It is because there is not any schools near their house have good quality. In many cases, this condition triggers many problems. For, students, they will spend more hours to travel from home to school. If this travel time is reduced, student will have more time to do other positive activities. The second problem in many cases, farther travel distance means higher transportation cost. So by reducing this distance, families may reduce their transportation expense. The third problem is longer travel distance may trigger traffic congestion because there will be more people are in the street for certain period of time.

Besides these good goals, this method still has several problems. One of these problems is there are some number of out of zone students which are students that are not accepted in school inside their zone. In some cases, there are not any schools inside their zone. In other cases, the students inside the zone of the school outnumber the school capacity. In other side, there are conditions that after the enrollment process is done, there are still empty seats in some schools. It is because the students in the zone outnumber the school capacity. This condition occurs because in many cities, the student location distribution is different with the school location distribution.

The other reason is that in the existing system, the enrollment process is done manually. It means that the students must choose their preferred schools in their application list. If they are fail to be accepted in these schools, they are fail in the entire enrollment process even there are available seats in their non preferred schools.

Based on this problem in this research, we propose new zone based school enrollment model. In this model, the process occurs automatically, so that, all students will be allocated as long as the total schools capacity in the enrollment system is enough. In this research, the model is developed by using k-means clustering as its basis method. The reason is that k-means method is one of common method in clustering process, especially in condition where the number of clusters has been defined before the clustering process occurs. Meanwhile, we modified the basic k-means method because in school enrollment system, there is limitation in school capacity so that the basic k-means clustering method cannot be implemented directly.

This research is the continuation from our previous research (Kusuma, 2018). In previous research, we developed enrollment models that combine the zone based method and student academic performance method. In our previous research, parameters that are evaluated are the success ratio, average student's national exam score and average student's home to school distance (Kusuma, 2018). In this research, we also evaluate the same parameters.

Problem in static zone based school enrollment system: One common method in the existing zone based school enrollment system is by using maximum distance. This method is adopted by many local governments in Indonesia such as in Bandung or Yogyakarta. In this method, the student is called as in zone student if the distance between his or her house and his or her school is lower than or equal to the defined maximum distance (r_{max}). This distance is various among cities. For example, government of Yogyakarta determines that the maximum distance is 5 km. Meanwhile, the government of Bandung determines that the maximum distance is 1.8 km. Some parameters that are used in this distance determination are the number of public schools, the distribution of the public schools and the distribution of the students. The illustration of this method is shown in Fig. 1.

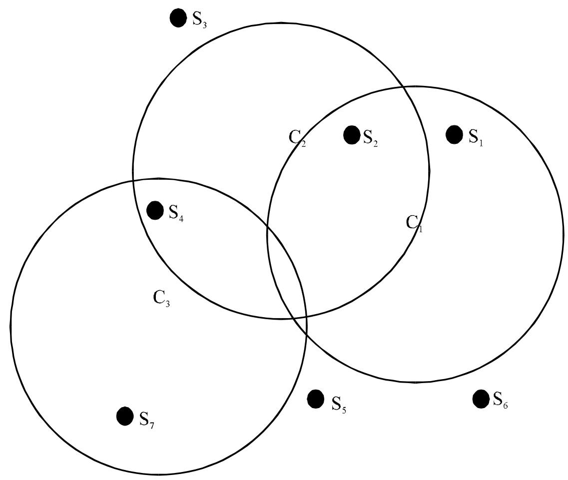

Fig. 1: School and student distribution illustration

Figure 1 shows there are three public schools that are involved in the school enrollment system that are run by local government (c_1-c_3). In the other side, there are seven students that apply to the school enrollment system (s_1-s_7). In this system, student has one school choice only. So, the relation between students and their preferred school is as follows $(<s_1, c_1>, <s_2, c_1>, <s_3, c_2>, <s_4, c_3>, <s_5, c_3>, <s_6, c_1>, <s_7, c_3>)$. Based on this scenario, students that are accepted in this system are s_1, s_2, s_4 and s_7. It is because these students are in the zone of their preferred school. In the other side, students that are rejected are s_3, s_5 and s_6. It is because these students are not in the zone of any schools in the system.

The good news is that in Indonesia, even the school must prioritize the in zone students, the out of zone students still can be accepted as long as there are still some available slots even all of in zone students that apply to enter this school has been accepted. The illustration is as follows. After all of in zone students have been accepted, the number of available seats is as follows

(2, 5, 0). Based on this available seat distribution, student s_3 and student s_6 are accepted in their preferred school. Unfortunately, student s_5 is still rejected because there is not any available seat in his preferred school.

MATERIALS AND METHODS

Proposed model: Based on this existing problem in this research, we propose new zone based model. In this model, the enrollment process runs automatically. It means that the student does not need to define his preferred school anymore. The system will allocate the student to the fittest school based on the distance between his home and the available school. This automatic based model is also implemented in our previous research (Kusuma, 2018). The goal of the automatic based model is to reduce the rejection potential because of the mismatch between the student's choice and the available school.

In this research, we use clustering method to allocate students to their school. In this model, the number of clusters (n_{cl}) is the number of schools (n_c) in the enrollment system. So, the number of cluster is defined previously before the clustering process runs. The goal of the clustering process is to minimize the total student's home to school distance. This concept is formalized by using Eq. 1-3:

$$g = \min(d_{tot}) \tag{1}$$

$$d_{tot} = \sum_{i=1}^{n_s} d\left(s_i, c_{j,i}\right) \tag{2}$$

$$d\left(s_i, c_{j,i}\right) = \left\|s_i - c_{j,i}\right\| \tag{3}$$

The explanation of these equations is as follows. Equation 1, the goal of the clustering process is minimizing the total distance. Equation 2 shows the total distance is the summation of all student's home to school distance. Equation 2 shows n_s denote the number of students. Variable s_i denotes student with index i. Variable $c_{j,i}$ denotes of school with index j and student i is belong to school j. Equation 3 shows the student's home to school distance is the euclidean distance between the student and his chosen school.

In this clustering process, the initial process is determining initial centroid position. Different with the basic k-means clustering method where the initial centroid position is determined randomly in this process, the initial centroid (e) position is the school position. This idea is formalized by using Eq. 4. In each iteration, node will be clustered to the centroid whose location is the nearest to the location of this centroid:

$$p\left(e_j, 0\right) = p\left(c_j\right) \tag{4}$$

Similar to the basic k-means clustering method, the next position of centroid j is the average position of the students that are belong to this cluster. This next centroid position determination is formalized by using Eq. 5 and 6. Equation 5 shows the $c_{j,\,x,\,n+1}$ denotes the next x position of centroid j. Equation 6 shws the $c_{j,\,y,\,n+1}$ denotes the next y position of centroid j. The $n_{m,\,j}$ is the number of students that belong to cluster j. $s_{i,\,x}$ is the x position of student i and $s_{i,\,y}$ denotes the y position of the student i:

$$c_{j,x,n+1} = \frac{\sum_{i=1}^{n_{m,j}} s_{i,x} \mid s \in S_j}{n_{m,j}} \tag{5}$$

$$c_{j,y,n+1} = \frac{\sum_{i=1}^{n_{m,j}} s_{i,y} \mid s \in S_j}{n_{m,j}} \tag{6}$$

In this clustering process, the system will iterates until the map is convergence. Convergence means that the current g value is equal to the previous g value. If the current g value is equal to the previous g value, the iteration will be stopped. Otherwise, the process will continue to iterate. This process is formalized by using Eq. 7:

$$A = \begin{cases} \text{Stop}, & g_n = g_{n-1} \\ \text{iterate}, & \text{else} \end{cases} \tag{7}$$

After the iteration stops, the next process is finalizing the members of the cluster. In the previous process, the school capacity or cluster capacity is ignored. In this finalization process, if the number of cluster members exceeds the cluster capacity, some members will be excluded from this cluster. In this process, the nearest students will be prioritized for being still in the cluster. For the excluded students, there are 2 possible conditions: transferred or rejected. This student will be transferred to other school as long as there is any available seat in other school. If there is not any available slot then this student will be rejected. This possible action is formalized by using Eq. 8. Equation 8 shows n_{tot_av} denotes the number of available slots in the system:

$$A = \begin{cases} \text{Transferred}, & n_{tor\_av} > 0 \\ \text{rejected}, & \text{else} \end{cases} \tag{8}$$

If the student is transferred, then the student will be transferred to other school that still has available slot and its location is the nearest to the student's location. This process is determined by using Eq. 9 and 10. Equation 9 shows C_{cand} denotes the school candidate set. n_{av} denotes the number of available slots. $c_{sel,i}$ denotes the school that accepts the student i:

$$C_{cand} = \{c / n_{av} > 0 \wedge c \in C\} \tag{9}$$

$$s_{sel,i} = c_i / \min\left(d\left(s_i, c_j\right)\right) \wedge c_j \in C_{cand} \tag{10}$$

RESULTS AND DISCUSSION

This model then is implemented into automatic school enrollment simulation application. This application is developed by using PHP language, so that, this application is a web based application. Besides this model, the static zone based method is also implemented into this application. After the simulation process, there are 2 results that can be compared to each other. This application is developed to observe the performance of the model.

In this application, the environment is a virtual city with its width is 25 km and its length is 25 km. In this city, there are some schools that its location is generated randomly. The school capacity is various among city. The capacity is generated randomly and it follows exponential distribution. Besides school, there are students that are generated in the environment and their location is generated randomly.

There are 4 scenarios in distributing students and schools in the environment. It is because in this research, we use 2 types of distribution: uniform and exponential. These scenarios characteristics illustration is shown in Fig. 2. When the system implements uniform distribution, the average value is the center of the virtual city. During the simulation, some variables are set as their default value. The default value of these variables is shown in Table 1.

This application then is used to evaluate the performance of the model. There are 3 testing groups. The first group is test with various numbers of colleges. The second group is test with various numbers of students. The third group is test with various numbers of average school capacities. In all tests, the observed variables are success ratio and average student's home to school distance. In this test, the maximum distance for the existing static method is set 5 km.

In the first test group, the number of schools ranges from 5-15 units. The step size is 1 unit. There are 5 simulation sessions for every step. Meanwhile, the other variables are set default. The success ratio result is shown in Table 2 and the average student's home to school distance is shown in Table 3.

Based on data in Table 2 it is shown that the increasing of the number of schools makes the success ratio increases too. It is because the increasing of the

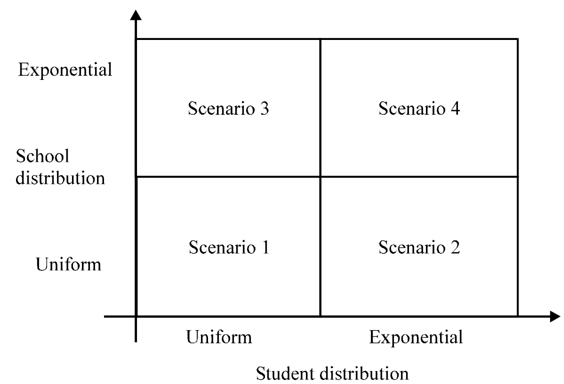

Fig. 2: Scenario characteristics

Table 1: Adjusted parameters default value

Parameter	Default value
n_c	10 units
n_s	1000 persons
c_{cap}	100 slots

Table 2: Success ratio result with various numbers of schools

	Success ratio (%)							
	Scenario 1 (%)		Scenario 2		Scenario 3		Scenario 4	
n_c (unit)	Proposed model	Static model	Proposed model	Static model	Proposed model	Static model	Proposed model	Static model
5	27	23	27	21	27	23	27	19
6	43	28	43	24	43	27	43	24
7	64	37	64	33	64	36	65	32
8	70	44	70	37	70	35	70	41
9	58	39	58	36	58	33	58	37
10	81	52	81	44	81	43	81	45
11	84	54	84	44	84	48	84	47
12	88	50	88	41	88	50	88	55
13	100	68	100	59	100	57	100	60
14	97	60	97	52	97	58	97	65
15	94	62	94	52	94	58	94	63

Table 3: Average student's home to school distance result with various numbers of schools

	Average student distance (km)							
	Scenario 1 (%)		Scenario 2		Scenario 3		Scenario 4	
n_c (unit)	Proposed model	Static model	Proposed model	Static model	Proposed model	Static model	Proposed model	Static model
5	4,671	3,115	5,173	3,250	4,542	3,142	5,835	3,031
6	5,681	2,977	6,896	3,142	5,637	3,100	7,546	2,946
7	6,541	3,085	8,223	3,101	7,852	3,100	7,839	2,899
8	6,444	3,079	7,735	3,152	8,280	2,963	7,907	2,848
9	5,551	3,023	6,450	3,077	7,375	3,013	7,299	2,890
10	5,512	2,965	6,787	3,073	7,268	3,010	7,746	2,712
11	6,073	2,973	7,476	3,035	6,885	2,986	7,324	2,732
12	6,715	2,974	7,587	2,990	6,537	2,965	5,805	2,708
13	5,446	2,946	6,487	2,977	6,749	2,907	6,696	2,705
14	5,827	2,863	6,784	2,883	6,658	2,853	5,716	2,746
15	5,553	2,778	6,631	2,860	6,563	2,915	6,015	2,654

Table 4: Success ratio result with various numbers of students

	Success ratio (%)							
	Scenario 1 (%)		Scenario 2		Scenario 3		Scenario 4	
n_s (persons)	Proposed model	Static model	Proposed model	Static model	Proposed model	Static model	Proposed model	Static model
500	100	61	100	49	100	49	100	56
600	100	58	100	47	100	47	100	53
700	94	53	94	44	94	43	94	51
800	94	52	94	38	94	46	94	56
900	93	55	93	47	93	48	93	49
1,000	80	48	80	35	80	43	80	44
1,100	91	53	91	47	91	43	91	51
1,200	79	47	79	40	79	37	79	44
1,300	66	50	66	40	66	40	66	44
1,400	56	40	56	36	56	35	56	40
1,500	67	41	67	39	67	42	67	44

number of schools makes the total capacity in the system increases too even there is not any change in capacity in every school. This condition occurs both in the proposed model and in the existing method. But the proposed model produces better success ratio rather than the existing method does. This condition occurs in all scenarios. When the system implements the proposed model, if the number of schools is same, the success ratio is same for all scenarios. Meanwhile, when the system implements the existing static method for same number of schools, there is variation in success ratio among scenarios even the gap is less significant.

Based on data in Table 3, the increasing of the number of schools does not affect the average student's home to school distance. When the number of schools increases, the average student's home to school distance tends to fluctuate with small deviation. This condition occurs both in the proposed model and in the existing static method. Meanwhile, the existing static model produces lower average student's home to school distance rather than the proposed model does. When the system implements the proposed model, the average student's home to school distance ranges from 4-7 km. When the system implements existing method, the average student's home to school distance ranges from 2-3 km.

Based on data in Table 3, the students and schools distributions affect the average student's home to school distance. When the system implements the proposed model, the average student's home to school distance in the first scenario is the lowest among the other scenarios. Meanwhile, the average student's home to school distance in the other 3 scenarios tends to similar. Meanwhile, when the system implements the existing static method, the average student's home to school distance is similar in all scenarios.

In the second test group, the number of students ranges from 500-1,500 persons. The step size is 100 persons. There are 5 simulation sessions for every step. Meanwhile, the other variables are set default. The success ratio result is shown in Table 4 and the average student's home to school distance is shown in Table 5.

Table 5: Average students' home to school distance result with various numbers of students

	Average students' distance (km)							
	Scenario 1 (%)		Scenario 2		Scenario 3		Scenario 4	
n_s (persons)	Proposed model	Static model	Proposed model	Static model	Proposed model	Static model	Proposed model	Static model
500	5,639	2,926	6,910	3,098	6,623	2,988	6,248	2,714
600	5,760	3,036	6,908	3,053	7,434	2,914	7,258	2,784
700	6,532	2,969	8,231	3,124	8,792	2,904	8,107	2,754
800	6,856	3,008	8,354	2,999	7,584	2,914	6,593	2,740
900	6,405	2,962	7,909	3,020	6,873	3,005	6,858	2,893
1,000	7,239	2,887	8,787	3,128	7,712	3,016	7,792	2,787
1,100	6,776	3,006	8,351	2,994	8,032	2,925	7,890	2,792
1,200	6,500	3,053	8,115	2,934	8,152	2,943	7,667	2,826
1,300	4,827	3,019	6,328	3,089	7,142	2,937	7,033	2,742
1,400	5,860	2,979	6,758	3,022	7,038	2,986	6,239	2,926
1,500	8,094	2,970	7,971	2,910	7,387	3,009	7,268	2,935

Table 6: Success ratio result with various average school capacities

	Success ratio (%)							
	Scenario 1 (%)		Scenario 2		Scenario 3		Scenario 4	
c_{cap} (slots)	Proposed model	Static model	Proposed model	Static model	Proposed model	Static model	Proposed model	Static model
50	45	35	45	32	45	32	45	34
60	55	40	55	38	55	32	55	38
70	65	47	65	40	65	32	65	42
80	75	45	75	39	75	44	75	40
90	75	46	75	41	75	42	75	47
100	93	56	93	51	93	49	93	54
110	89	55	89	45	89	42	89	47
120	84	48	84	46	84	46	84	52
130	94	54	94	46	94	42	94	48
140	97	55	97	43	97	52	97	52
150	91	51	91	46	91	48	91	49

Based on data in Table 4, it is shown that the increasing of the number of students makes the success ratio decreases. This condition occurs both when the system implements the proposed model or the existing static method. But there is difference between these models. The success ratio drops faster when the system implements the proposed model. When the number of students is 500 persons, the success ratio is 100 %. When the number of students is 1,500 persons, the success ratio is 67%. This condition occurs in all scenarios. So, it can be said that when the system implements the proposed model, the student and school locations distribution does not affect the success ratio.

The success ratio drops slower when the system implements the existing static method due to the increasing of the number of students. The success ratio ranges from 49-61% when the number of students is 500 persons. The success ratio ranges from 39-44% when the number of students is 1,500 persons. By comparing all among scenarios, there is not any significant difference in success ratio. So, it can be said that when the system implements the existing static method, even there is variation, the student and school locations distribution does not affect the success ratio.

Based on data in Table 5, it is shown that the increasing number of students does not affect the average student's home to school distance. This condition occurs when the system implements the proposed model or the existing static method. Meanwhile, the existing static method produces lower average student's home to school distance rather than the proposed model does.

During the increasing of the number of students, the average student's home to school distance tends to fluctuate in all scenarios. But there is difference between models. When the system implements the existing static method, the variation is not significant. The difference between scenarios is also less significant. When the system implements the proposed model, the variation of the average student's home to school distance due to the increasing of the number of students is significant. By comparing among scenarios, the lowest average student's home to school distance occurs in the first scenario. Meanwhile, the highest average student's home to school distance occurs in the second scenario. So, it can be said that when the system implements the proposed model, the student and school location distribution affect the average student's home to school distance.

Table 7: Average Student's home to school distance result with various average school capacities

	Average students' distance (km)							
	Scenario 1 (%)		Scenario 2		Scenario 3		Scenario 4	
c_{cap} (slot)	Proposed model	Static model	Proposed model	Static model	Proposed model	Static model	Proposed model	Static model
50	5,073	2,963	5,930	3,059	7,919	2,972	6,494	2,903
60	6,090	3,047	6,497	3,136	8,165	2,910	7,585	2,723
70	5,933	3,037	6,515	3,085	9,027	2,867	7,932	2,652
80	7,544	2,981	8,709	3,028	8,027	3,087	8,358	2,822
90	6,850	2,959	7,518	2,995	7,312	2,967	7,102	2,796
100	6,366	2,938	7,159	2,964	7,542	2,868	7,101	2,763
110	6,161	2,976	7,316	3,082	7,710	2,987	7,460	2,654
120	7,083	2,962	7,625	2,982	7,768	2,962	7,153	2,858
130	5,867	3,058	7,180	3,080	9,405	2,913	8,880	2,744
140	7,222	2,966	8,477	3,057	7,190	3,023	7,421	2,834
150	6,905	2,915	7,763	2,941	7,983	2,957	7,528	2,760

In the third test group, the average school capacity ranges from 50 persons to 150 slots per school. The step size is 10 slots. There are 5 simulation sessions for every step. Meanwhile, the other variables are set default. The success ratio result is shown in Table 6 and the average student's home to school distance is shown in Table 7.

Based on data in Table 6, it is shown that increasing of the school capacity makes the success ratio increase. This condition occurs both in the proposed mode and the existing static method. But the proposed model produces higher success ratio rather than the existing static method. Based on this result, due to the increasing of the school capacity, the success ratio grows faster when the system implements the proposed model rather than when the system implements the existing static method. When the system implements the proposed model, the change in success ratio is as follows. The success ratio is 45% when the average school capacity is 50 slots. The success ratio is 91% when the average school capacity is 150 slots. This condition occurs in all scenarios. It can be said that when the system implements the proposed model, due to the change in average school capacity, the student and school locations distribution does not affect the success ratio.

When the system implements the existing static method, the change in success ratio is as follows. The success ratio ranges from 32-35% when the average school capacity is 50 slots. The success ratio ranges from 46- 51% when the average school capacity is 150 slots. By comparing among scenarios, the difference of the success ratio when the average school capacity is same is not significant. It can be said that when the system implements the existing static method, the student and school location distribution does not affect the success ratio due to the change in average school's capacity.

Based on data in Table 7, it is shown that the increasing of school capacity does not affect the average student's home to school distance. The average student's home to school distance tends to fluctuate due to change of the average school capacity. This condition occurs both in the proposed model and in the existing static method. Meanwhile, the existing method produces lower average student's home to school distance rather than the proposed model does.

The student's home to school distance ranges from 5-9 km when the system implements the proposed model. By comparing among scenarios in the first scenario, the average student's home to school distance is in the lowest value. Meanwhile, in the third scenario, the average student's home to school distance is in the highest value. So, it can be said that when the system implements the proposed model, the student and school location distribution affects the average student's home to school distance.

The student's home to school distance ranges from 2-3 km when the system implements the existing static method. By comparing among scenarios, there is not any significant difference in average student's home to school distance. So, it can be said that when the system implements the existing static method, due to change in school capacity, the student and school locations distribution does not affect the average student's home to school distance.

CONCLUSION

Based on the explanation above, the proposed model has been developed and has been implemented into school enrollment simulation application. This proposed model has been also compared with the existing static maximum distance zone based method. Based on the test, there research findings in the relation between the adjusted variables and the observed variables.

The adjusted variables affect the success ratio. The number of schools and the average school capacities have positive correlation with the success ratio. In the other side, the number of students has negative correlation with the success ratio. Meanwhile, the proposed model produces higher success ratio rather than the existing static method. The student and school location distribution does not affect the success ratio.

In the other side, the adjusted variables do not affect the student's home to school distance. Due to change in adjusted variables, the student's home to schools tends to fluctuate. But the existing static method produces lower average student's home to school distance rather than the proposed model does. The student and school location distribution affect less significant in average student's home to school distance when the system implements the proposed model. In the other side, the student and school location does not affect the average student's home to school distance when the system implements the existing static method.

RECOMMENDATIONS

There are many future research potentials in developing school enrollment system as the continuation of this research. In this research, the system has been tested by using simulation data. By testing this system with real data will benefits in finding the advantages and the limitations of this proposed model. In the other side with the decentralization concept in Indonesia, the local government can make improvisation in the implementation of the central policy. So, study in comparing this proposed model in many real local data will be interesting.

ACKNOWLEDGEMENT

We would like to thank to Telkom University for its support in this work, especially in funding and research facilities. We also thank to our colleague, Dr. Tito Waluyo Purboyo for giving better and deeper insight about machine learning, especially, in clustering process.

REFERENCES

Banks, G., 2005. Comparing school systems across Australia. Proceedings of the International Conference on Address to ANZSOG: Schooling in the 21st Century: Unlocking Human Potential Vol. 21, September 28-29, 2005, Sydney, Australia, pp: 1-25.

Beaven, C., 2003. Parental choice or school choice: Who benefits from the removal of zoning?'. New Zealand Ann. Rev. Educ., 12: 111-126.

Boser, U., 2013. Size matters: A look at school-district consolidation. Master Thesis, Center for American Progress, Washington, DC., USA.

Devonport, A., 2017. The impact of secondary school enrolment schemes on school desirability, academic achievement and transport. Master Thesis, Departement of Geography, University of Canterbury, Christchurch, New Zealand.

Dewi, S.W. and C. Elyda, 2013. City applies zoning system to school enrollment. The Jakarta Post, Jakarta, Indonesia. https://www.thejakartapost.com/news/2013/06/11/city-applies-zoning-system-school-enrollment.html

Fossey, R., 1994. Open enrollment in Massachusetts: Why families choose. Educ. Eval. Policy Anal., 16: 320-334.

Hastings, J.S., R. Van Weelden and J. Weinstein, 2007. Preferences, information and parental choice behavior in public school choice. Master Thesis, Department of Economics, Yale University, New Haven, Connecticut.

Hastings, J.S., T.J. Kane and D.O. Staiger, 2005. Parental preferences and school competition: Evidence from a public school choice program (No. w11805). Master Thesis, National Bureau of Economic Research, Cambridge, Massachusetts.

Hill, P.T., 1990. High Schools with Character. RAND Corporation, New York, USA., ISBN:9789990690491, Pages: 97.

Kalata, K. and J. McEwen, 2015. Chicago: A Achoice District. Illinois Network of Charter Schools, Chicago, Illinois,.

Kusuma, P.D., 2018. Distance-score combined model in automatic high school student enrollment system. J. Theor. Appl. Inf. Technol., 96: 6699-6713.