Research Article

K-Means Method for Clustering Average Length of School (ALS) in Central Java

Ajeng Tiara Wulandari*, Jati Sumarah

[1]Politeknik Dharma Patria, Kebumen, Indonesia

Abstract.
Average School Length (ALS) represents the level of performance of every person in a school region. The more years of education, the better the education obtained by the population; therefore, it becomes more essential as it can demonstrate the quality of the human resources in a given region. In addition, extensive research implies that the average length of schooling influences economic growth considerably. If the average length of education improves, then the number of unemployed and poor in an area decreases and affects economic growth positively and significantly. The study's goal was to carry out an investigation utilizing artificial intelligence in the form of a cluster mapping on the averages of regeneration in central Java. This needs to be done to get a macro view of the level of development of the average school years through regional mapping over the last few years. The data set used can be found on the website of the Central Java Provincial Statistics Agency, which is the subject of the 2017-2019 average school year by sex. The solution approach is k-means, which is part of the data collection process. High and low clusters are the number used in this investigation. Prior to the k-means approach, pre-processing is conducted out by taking from 2017-2019 the average RLS number based on gender. The results of the computed average value were analyzed with k-means. The study found that out of 35 provinces, eight (23%) provinces were of the high cluster (cluster 1) and 27 of the low cluster (cluster 0) (77%). More than 70% of Central Java regions have remained common ALS, according to results.

Keywords: K-Means, clustering, ALS

Corresponding Author: Ajeng Tiara Wulandari; email: ajeng-tiarawulandaridp@gmail.com

Published 26 May 2023

Publishing services provided by Knowledge E

Selection and Peer-review under the responsibility of the ICASI Conference Committee.

1. Introduction

Average Length of School (ALS) reflects the average number of years that inhabitants have spent completing formal education at all levels of formal education that they have pursued This figure can also be used to describe the level of education received by the local population [1]. This figure is calculated using three factors that are all present at the same time: the variable relating to school involvement, current/previous level/class, and the last diploma held. The population calculated in this average length of schooling is the population aged 15 years and over. The greater the number of years spent in school, the higher the population's level of education, so this indicator is very important because it can show the quality of human resources in an area. The length of schooling also

has a big impact on the economy, according to many studies that have looked at this question [2]. Aforementionedmeans, a rise in the ALS, less people will be unemployed and poor in a given location. It will certainly have a positive and significant influence on economic growth [3]. In this proposed paper, the average length of schooling will be discussed specifically in the districts/cities in the province of Central Java-based on gender. As is known, After West Java and East Java, Central Java has the third-highest population. Moreover, based on data from the Indonesian Central Statistics Agency in Semester 1 (March) 2020, Central Java is the sixth province that has the highest percentage of the urban poor with 10.09% after West Nusa Tenggara (14.90%), Bengkulu (14.77%), South Sumatra (12.16%), DI Yogyakarta (11.53) and Jambi (10.41%). Meanwhile, the average length of schooling in Indonesia is still based on data from the Indonesian Central Statistics Agency, in 2019 Central Java was in the fifth lowest (7.53) after the provinces of Papua (6.65), West Nusa Tenggara (7.27), West Kalimantan (7.31) and West Papua (7.44) [4]. The third-most populous province in Java, Central Java, despite its location on the island, which should be able to be an example for other provinces in Indonesia, it is very sad because, in fact, the average length of schooling in this province is very low.

Therefore, grouping the Average Years of Schooling is very important, as information and a special government barometer for local governments in each district and city in Central Java Province in determining related policies in the field of education. The specific purpose of this research is to provide input and information for the Central Java provincial government to maximize efforts and care to improve areas that have a low Average Years of Schooling and maintain a stable Average Years of Schooling for those areas. Areas where the ALS is already high. The research dataset is from data on ALS by Gender in Central Java Province from 2017 to 2019, consisting of 29 Regencies and 6 Cities obtained from the Central Java Statistics Agency. The clustering method used in this study is the K-Means Clustering data mining algorithm. Because data mining is an algorithm that is widely used to deal with data classification problems [5]–[9], as well as data clustering [10]–[14].

Numerous past research on the K-Means technique is used to group data exist, including the following: A study was conducted to categorize disaster-prone locations in Indonesia based on provinces. The results of this study are in the form of grouping data on locations prone to natural disasters which are categorized into three groups, specifically, four provinces are included in the high-density cluster, the conventional

Table 1: Average Length of School (ALS) by Gender (Years).

No	Central Java Region	Man			Woman		
		2019	2018	2017	2019	2018	2017
1	Cilacap District	7.42	7.41	7.4	6.47	6.45	6.44
2	Banyumas District	7.95	7.94	7.93	7.14	7.12	7.11
3	Purbalingga District	7.47	7.43	7.28	6.76	6.59	6.48
4	Banjarnegara District	6.66	6.58	6.57	6.32	6.09	6.08
5	Kebumen District	7.97	7.94	7.9	7.09	6.87	6.82
6	Purworejo District	8.64	8.43	8.42	7.49	7.34	7.33
7	Wonosobo District	7.01	7	6.76	6.51	6.5	6.27
8	Magelang District	8.3	8	7.87	7.28	7.17	6.95
9	Boyolali District	8.46	8.45	8.44	6.94	6.85	6.72
10	Klaten District	8.94	8.93	8.92	7.77	7.65	7.62
11	Sukoharjo District	9.72	9.48	9.47	8.67	8.41	8.25
12	Wonogiri District	7.61	7.52	7.39	6.71	6.49	6.13
13	Karanganyar District	9.32	9.31	9.3	7.93	7.91	7.74
14	Sragen District	8.12	7.89	7.84	6.61	6.6	6.3
15	Grobogan District	7.35	7.19	7.18	6.34	6.14	6.13
16	Blora District	7.14	6.99	6.98	6.06	5.96	5.95
17	Rembang District	7.71	7.53	7.52	6.61	6.41	6.4
18	Pati District	7.78	7.77	7.51	6.69	6.64	6.57
19	Kudus District	9.2	9.19	8.93	8.1	8.09	7.7
20	Jepara District	7.89	7.81	7.73	6.81	6.8	6.67
21	Demak District	8.27	8.26	8.25	6.91	6.82	6.81
22	Semarang District	8.44	8.36	8.35	7.62	7.44	7.43
23	Temanggung District	7.52	7.2	7.12	6.8	6.72	6.71
24	Kendal District	7.91	7.64	7.42	6.63	6.47	6.3
25	Batang District	7.33	7.32	7.31	6.17	6.16	6.05
26	Pekalongan District	7.21	7.07	7.06	6.56	6.31	6.15
27	Pemalang District	6.94	6.9	6.89	5.92	5.75	5.65
28	Tegal District	7.44	7.3	7.15	6.31	6.09	5.97
29	Brebes District	6.76	6.75	6.74	5.38	5.37	5.36
30	Magelang City	11.05	10.93	10.92	9.96	9.95	9.94
31	Surakarta City	11.1	11.09	10.96	10.09	10.01	9.84
32	Salatiga City	10.99	10.91	10.68	9.94	9.93	9.57
33	Semarang City	11.41	11.29	11.28	10.15	10.14	10.13
34	Pekalongan City	8.95	8.94	8.93	8.5	8.22	8.21
35	Tegal City	8.7	8.69	8.68	7.82	7.71	7.45

Source: Central Java Provincial Statistics Agency [15]

cluster consists of fourteen provinces, and there are sixteen provinces in the bottom cluster [16]. The following research was undertaken to categorize population density,

human development index, open unemployment rate, and school enrollment rate by Indonesian provinces. The study's findings indicate that cluster 1 contains 12 provinces, cluster 2 contains six provinces, cluster 3 contains one province, cluster 4 contains six provinces, and cluster 5 contains nine provinces [17]. The next research was conducted to cluster the distribution of rabies cases in the city of Palembang using K-Means data mining. Data processing in this study using RapidMiner software with the result that from 16 sub-districts in Palembang, seven sub-districts are included in the very rabies-prone area cluster (C0), while four sub-districts are included in the rabies-prone area cluster (C1), and five sub-districts are included in the regional cluster. not susceptible to rabies (C2) [18]. These related studies are the background for researching to classify the ALS in Regencies and Cities in Central Java Province.

2. Methods

2.1. Research Data

The research dataset is in the form of data on ALS by Gender in Central Java Province from 2017 to 2019, which consists of 29 Regencies and 6 Cities obtained from the Central Java Statistics Agency (Table 1) [15]. The clustering method used in this study is the K-Means Clustering data mining algorithm.

2.2. K-Means Stages

The flow chart of the K-Means Clustering method can be seen in Figure 1.

Figure 1's steps can be broken down as follows: Calculate the total number of clusters (k). The next step is to calculate the center's value (Centroid). After then, the initial stage of determining the centroid value was done at random. Then, for each record, compute the distance to the centroid. Euclidean, Manhattan/City Block, and Minkowsky distances can all be used to measure the distance of the data to the group's center. There are benefits and drawbacks to each approach. Euclidean Distance is chosen as the centroid distance in this investigation. Put items in groups according to how far they are from the nearest Centroid. Repeat steps 3 through 4 until the Centroid reaches its ideal value, and then stop.

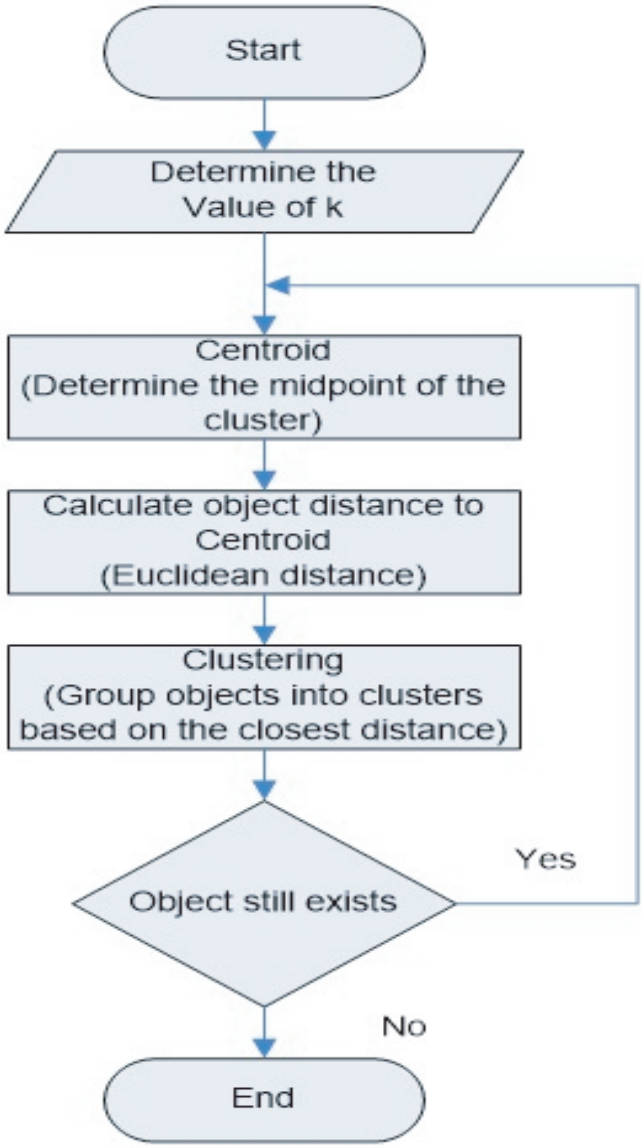

Figure 1: K-Means Stages.

3. Results and Discussion

The following are the results of the sample dataset that will be used in grouping the average percentage of school years using the K-Means method:

The dataset in Table 2 will be processed using the RapidMiner software using the k-means method. In this study, the mapping process carried out will produce two clusters, namely the high cluster to the old average (C1) and the low cluster to the old average (C2). Determine the centroid in iteration 1 taken from the 29th data and 33rd data;

Iteration 1

Table 2: Processed Average Length of School (ALS) percentage dataset.

No	Central Java Region	Man	Woman
1	Cilacap District	7.41	6.45
2	Banyumas District	7.94	7.12
3	Purbalingga District	7.39	6.61
4	Banjarnegara District	6.60	6.16
5	Kebumen District	7.94	6.93
6	Purworejo District	8.50	7.39
7	Wonosobo District	6.92	6.43
8	Magelang District	8.06	7.13
9	Boyolali District	8.45	6.84
10	Klaten District	8.93	7.68
11	Sukoharjo District	9.56	8.44
12	Wonogiri District	7.51	6.44
13	Karanganyar District	9.31	7.86
14	Sragen District	7.95	6.50
15	Grobogan District	7.24	6.20
16	Blora District	7.04	5.99
17	Rembang District	7.59	6.47
18	Pati District	7.69	6.63
19	Kudus District	9.11	7.96
20	Jepara District	7.81	6.76
21	Demak District	8.26	6.85
22	Semarang District	8.38	7.50
23	Temanggung District	7.28	6.74
24	Kendal District	7.66	6.47
25	Batang District	7.32	6.13
26	Pekalongan District	7.11	6.34
27	Pemalang District	6.91	5.77
28	Tegal District	7.30	6.12
29	Brebes District	6.75	5.37
30	Magelang City	10.97	9.95
31	Surakarta City	11.05	9.98
32	Salatiga City	10.86	9.81
33	Semarang City	11.33	10.14
34	Pekalongan City	8.94	8.31
35	Tegal City	8.69	7.66

Table 3: Centroid Data Iteration 1.

Atribut	Man	Woman
C1	6,75	5,37
C2	11,33	10,14

$c\,(1,1) = \sqrt{(7,41-6,75)^2+(6,45-5,37)^2} = 1,83$, and so on until you get:
$c\,(35,2) = \sqrt{(8,69-11,33)^2+(7,66-10,14)^2} = 8,79$

TABLE 4: Cluster Center Distance Calculation Results 1.

No	Central Java Region	C1	C2	Nearest Distance	Data Clustering
1	Cilacap District	1,83	17,51	1,83	C1
2	Banyumas District	4,26	12,49	4,26	C1
3	Purbalingga District	2,18	16,39	2,18	C1
4	Banjarnegara District	0,78	20,54	0,78	C1
5	Kebumen District	3,61	13,72	3,61	C1
6	Purworejo District	5,81	10,41	5,81	C1
7	Wonosobo District	1,29	18,19	1,29	C1
8	Magelang District	4,42	12,31	4,42	C1
9	Boyolali District	3,85	13,79	3,85	C1
10	Klaten District	7,52	8,45	7,52	C1
11	Sukoharjo District	12,25	4,65	4,65	C2
12	Wonogiri District	1,91	17,49	1,91	C1
13	Karanganyar District	8,76	7,22	7,22	C2
14	Sragen District	2,48	16,60	2,48	C1
15	Grobogan District	1,18	19,58	1,18	C1
16	Blora District	0,67	21,51	0,67	C1
17	Rembang District	2,05	17,18	2,05	C1
18	Pati District	2,53	15,94	2,53	C1
19	Kudus District	9,08	6,96	6,96	C2
20	Jepara District	2,99	14,94	2,99	C1
21	Demak District	3,69	13,91	3,69	C1
22	Semarang District	6,16	9,93	6,16	C1
23	Temanggung District	2,42	15,58	2,42	C1
24	Kendal District	2,11	17,16	2,11	C1
25	Batang District	1,14	20,11	1,14	C1
26	Pekalongan District	1,30	18,65	1,30	C1
27	Pemalang District	0,32	23,48	0,32	C1
28	Tegal District	1,11	20,16	1,11	C1
29	Brebes District	0,00	27,33	0,00	C1
30	Magelang City	25,19	0,40	0,40	C2
31	Surakarta City	25,55	0,30	0,30	C2
32	Salatiga City	23,85	0,57	0,57	C2
33	Semarang City	27,33	0,00	0,00	C2
34	Pekalongan City	10,83	5,74	5,74	C2
35	Tegal City	7,18	8,79	7,18	C1

Iteration 2

Table 5: Centroid Data Iteration 2.

Atribut	Man	Woman
C1	7,65	6,62
C2	10,14	9,06

$c\,(1,1) = \sqrt{(7,41-7,65)^2+(6,45-6,62)^2} = 0,27$, and so on until you get:

$$c\,(35,2) = \sqrt{(8,69-10,14)^2+(7,66-9,06)^2} = 3,4$$

Table 6 shows that from 35 provinces, 8 provinces (23 percent) are in the high cluster (cluster_1) and 27 provinces (77 percent) are in the low cluster (cluster_0). The high clusters are Sukoharjo Regency, Karanganyar Regency, Kudus Regency, Magelang City, Surakarta City, Salatiga City, Semarang City, Pekalongan City. While the low clusters are Cilacap Regency, Banyumas Regency, Purbalingga Regency, Banjarnegara Regency, Kebumen Regency, Purworejo Regency, Wonosobo Regency, Magelang Regency, Boyolali Regency, Klaten Regency, Wonogiri Regency, Sragen Regency, Grobogan Regency, Blora Regency, Rembang Regency, Regency Pati, Jepara Regency, Demak Regency, Semarang Regency, Temanggung Regency, Kendal Regency, Batang Regency, Pekalongan Regency, Pemalang Regency, Tegal Regency, Brebes Regency, Tegal City.

4. Conclusions

Based on the results of the study, it can be concluded that the application of the k-means method can be carried out in the form of clusters in the ALS area in the province of Central Java. The results obtained indicate that over 70% of areas in Central Java still have low ALS.

Acknowledgements

Our thanks go to Direktorat Riset dan Pengabdian Masyarakat, Deputi Bidang Penguatan Riset and Pengembangan, Kementerian Riset and Teknologi/ Badan Riset and Inovasi Nasional (Ristek/BRIN) who has provided novice lecturer research grants (PDP) in 2020 for 2021 funding. Furthermore, thank you to the chairman and staff of the Dharma Patria Polytechnic LPPM who have facilitated PDP activities starting from the preparation of proposals to research reports.

Table 6: Calculation Result of Center Distance Cluster Iteration 2.

No	Central Java Region	C1	C2	Nearest Distance	Data Clustering
1	Cilacap District	0,27	9,51	0,27	C1
2	Banyumas District	0,54	5,94	0,54	C1
3	Purbalingga District	0,26	8,74	0,26	C1
4	Banjarnegara District	1,26	11,91	1,26	C1
5	Kebumen District	0,38	6,74	0,38	C1
6	Purworejo District	1,43	4,43	1,43	C1
7	Wonosobo District	0,77	10,14	0,77	C1
8	Magelang District	0,67	5,79	0,67	C1
9	Boyolali District	0,85	6,62	0,85	C1
10	Klaten District	2,40	3,11	2,40	C1
11	Sukoharjo District	5,24	0,96	0,96	C2
12	Wonogiri District	0,18	9,47	0,18	C1
13	Karanganyar District	3,20	2,26	2,26	C2
14	Sragen District	0,31	8,71	0,31	C1
15	Grobogan District	0,58	11,05	0,58	C1
16	Blora District	1,01	12,51	1,01	C1
17	Rembang District	0,09	9,23	0,09	C1
18	Pati District	0,03	8,33	0,03	C1
19	Kudus District	3,26	2,23	2,23	C2
20	Jepara District	0,18	7,61	0,18	C1
21	Demak District	0,66	6,77	0,66	C1
22	Semarang District	1,50	4,19	1,50	C1
23	Temanggung District	0,39	8,21	0,39	C1
24	Kendal District	0,03	9,20	0,03	C1
25	Batang District	0,57	11,41	0,57	C1
26	Pekalongan District	0,62	10,41	0,62	C1
27	Pemalang District	1,46	14,02	1,46	C1
28	Tegal District	0,60	11,45	0,60	C1
29	Brebes District	2,46	16,99	2,46	C1
30	Magelang City	14,41	1,62	1,62	C2
31	Surakarta City	14,70	1,76	1,76	C2
32	Salatiga City	13,42	1,29	1,29	C2
33	Semarang City	16,08	2,36	2,36	C2
34	Pekalongan City	4,15	1,76	1,76	C2
35	Tegal City	2,12	3,40	2,12	C1

References

[1] Maulana H, et al. Penyesuaian Sistem Tata Kelola Pada Institut Teknologi Kalimantan Dengan Menggunakan Cobit 2019 Pelayanan Teknis (UPT) dan tim ad hoc. Salah

satu tugas pengelola TI di perguruan tata kelola teknologi informasi adalah Control Objective for Information a. *J Sist Inf*. 2020;12(2):2060–2074.

[2] Wabang K, Rahma Y, Widodo AP, et al. "Tata Kelola Teknologi Informasi Menggunakan Cobit 2019 Pada Psi Universitas Muria Kudus." *Jurnal Teknol dan* 2021;VII(3):275–282.

[3] Sugiyono. "Metode Penelitian Bisnis (Pendekatan Kuantitatif, Kualitatif, Kombinasi dan R&D)." in *Metodelogi Penelitian*, 2017.

[4] D. Steuperaert, "COBIT 2019: A significant update." *EDPACS*, https://doi.org/10.1080/07366981.2019.1578474.

[5] ISACA. COBIT 2019 Introduction and Methodology. ISACA; 2019.

[6] Kasma VS, Sutikno S, Surendro K. "Design of e-Goverment security governance system using Cobit 2019." *Int Conf ICT Smart Soc.* 2019.

[7] Fikri AM, Priastika HS, Octaraisya N, Sadriansyah S, Trinawati LH. "Rancangan Tata Kelola Teknologi Informasi Menggunakan Framework COBIT 2019 (Studi Kasus: PT XYZ)." *Inf Manag Educ Prof J Inf Manag.* https://doi.org/10.51211/imbi.v5i1.1410.

[8] Gusni RS, Kraugusteeliana K, Pradnyana IW. "Analisis Tata Kelola Keamanan Sistem Informasi Rumah Sakit XYZ Menggunakan Cobit 2019 (Studi Kasus pada Rumah Sakit XYZ)." in Konferensi Nasional Ilmu Komputer (KONIK) 2021; 2021.

[9] Riantini FI, Mulyana DI. Implementasi Tata Kelola Teknologi Informasi Menggunakan Framework Cobit 5 Pada Direktorat Jenderal Bea Dan Cukai. Cki Spot. 2019;12(1):1–14.

[10] Made IG, Dharma S, Made IG, Sasmita A, Suwija IM. "Evaluasi Dan Implementasi Tata Kelola Ti Menggunakan Cobit 2019 (Studi Kasus Pada Dinas Kependudukan Dan Pencatatan Sipil Kabupaten Tabanan)." *J Ilm Teknol dan Komput.* 2021;2(2).

[11] ISACA. *COBIT 2019 Framework Introduction and methodology*. 2019.

[12] Safitri A, Syafii I, Adi K. Measuring the performance of information system governance using framework COBIT 2019. *Int J Comput Appl*. 2021;174(31):23–30.

[13] Atrinawati LH, Ramadhani E, Fiqar TP, Wiranti YT, Abdullah AINF, Saputra HMJ, et al. Assessment of process capability level in university XYZ based on COBIT 2019. *J Phys*. Conference Series; https://doi.org/10.1088/1742-6596/1803/1/012033

[14] Gerl A, Von Der Heyde M, Groß R, Seck R, Watkowski L. "Applying COBIT 2019 to IT Governance in Higher Education Establishing IT governance for the collaboration of all universities and universities of applied sciences in Bavaria." *Informatik 2020*, 2020.

[15] Belo GI, Wiranti YT, Atrinawati LH. “Perancangan Tata Kelola Teknologi Informasi Menggunakan Cobit 2019 Pada PT Telekomunikasi Indonesia Regional VI Kalimantan.” *Jusikom Prima*. *Jurnal Sist Inf Ilmu Komput. Prima*; 2020.

[16] Gerl A, Von Der Heyde M, Groß R, Seck R, Watkowski L. “Applying COBIT 2019 to IT Governance in Higher Education.” *Inform. 2020 - Gesellschaft für Inform.*, 2021.

[17] Ishlahuddin A, Handayani PW, Hammi K, Azzahro F. “Analysing IT governance maturity level using COBIT 2019 framework: A case study of small size higher education institute (XYZ-edu).” in *2020 3rd International Conference on Computer and Informatics Engineering, IC2IE 2020*, https://doi.org/10.1109/IC2IE50715.2020.9274599.

[18] Wardani S, Puspitasari M. Audit Tata Kelola Teknologi Informasi Mengunakan Framework Cobit Dengan Model Maturity Level (Studi Kasus Fakultas Abc). *J Teknol*. 2014.

[19] Nagel RG, Pasyha G, Wagiu EB. “Analisa Tata Kelola Sistem Informasi Universitas Advent Indonesia menggunakan Framework COBIT 5 (Domain EDM).” *Semin Nas Inform.* 2020.

[20] White SK. What is COBIT? A framework for alignment and governance. CIO (Framingham Mass). 2019.

[21] Nachrowi E, Nurhadryani Y, Sukoco H. “Evaluation of governance and management of information technology services using Cobit 2019 and ITIL 4.” *J RESTI (Rekayasa Sist dan Teknol Informasi)*, 2020.