

DOI <http://dx.doi.org/10.36722/sst.v7i1.877>

Simulasi Rancangan Pemetaan Sekolah dengan Metode Algoritma *Machine Learning* Menggunakan *Software RapidMiner*

Aisyah Sabrina Aprilia¹, Budi Aribowo¹, Ahmad Chirzun¹

¹Program studi Teknik Industri, Fakultas Sains dan Teknologi, Universitas Al Azhar Indonesia, Komplek Masjid Agung Al Azhar, Jalan Sisingamangaraja, Kebayoran Baru, Jakarta Selatan, 12110

Penulis untuk Korespondensi/E-mail: aisyahsabrinaaprilia@gmail.com

Abstract – The growth in the number of educational institutions creates competition which encourages each educational institution to have a special strategy to deal with it. The challenge that must also be faced in the world of education is the uneven quality of education in various regions. One way to face the challenge of equal distribution of school quality is to create programs that suit the needs of each school. In this study, solving the problem of mapping schools owned by the XYZ educational foundation was carried out. The results of this study obtained a design of a school mapping indicator instrument to assess the quality of each school. Then obtained a simulation of school mapping design from the results of unsupervised learning with the K-Means and K-Medoids Clustering methods, as well as a simulation of predicting school mapping patterns from the results of supervised learning with the Decision Tree C4.5 method. The results of K-Medoids were selected for the proposed school mapping with a Davies Bouldin index value of 0.112. The model cluster owned by K-Medoids, namely Excellent School, has 44 schools; at Good School, there are 36 schools; and in the Improvement School, there are 20 schools. Meanwhile, the prediction pattern with Decision Tree C4.5 obtained rules with the dominant indicator attributes (IND) 1, 2, 3, and 4. Also, the prediction simulation results using the 80:20 ratio decision tree model show the new testing data with the assumption that the 101st school goes to cluster_1 with an accuracy rate of 95%.

Abstrak – Pertumbuhan jumlah lembaga pendidikan menimbulkan persaingan yang mendorong tiap lembaga pendidikan harus memiliki strategi khusus untuk menghadapinya. Tantangan yang juga harus dihadapi dalam dunia pendidikan adalah kualitas pendidikan yang tidak merata di berbagai daerah. Salah satu cara untuk menghadapi tantangan pemerataan kualitas sekolah adalah dengan membuat program yang sesuai dengan kebutuhan masing-masing sekolah. Pada penelitian ini, dilakukan pemecahan masalah pemetaan sekolah yang dimiliki yayasan pendidikan XYZ. Hasil dari penelitian ini didapatkan suatu rancangan instrumen indikator pemetaan sekolah untuk melakukan penilaian terhadap kualitas masing-masing sekolah. Kemudian didapatkan simulasi rancangan pemetaan sekolah dari hasil *unsupervised learning* dengan metode *K-Means* dan *K-Medoids Clustering*, serta simulasi prediksi pola pemetaan sekolah dari hasil *supervised learning* dengan metode *Decision Tree C4.5*. Hasil *K-Medoids* terpilih untuk usulan pemetaan sekolah dengan nilai *davies bouldin index* 0.112. *Cluster model* yang dimiliki *K-Medoids*, yaitu *Excellent School* terdapat 44 sekolah; pada *Good School* terdapat 36 sekolah; dan pada *Improvement School* terdapat 20 sekolah. Sedangkan, pola prediksi dengan *Decision Tree C4.5* didapatkan *rules* dengan atribut indikator (IND) 1, 2, 3, dan 4 yang mendominasi. Serta, hasil simulasi prediksi dengan menggunakan model *decision tree* rasio 80:20 didapatkan bahwa data *testing* baru dengan asumsi sekolah ke-101 masuk ke *cluster_1* dengan tingkat akurasi sebesar 95%.

Keywords - *decision tree, K-means, K-medoids, school mapping indicator instrument.*

PENDAHULUAN

Pendidikan menjadi hal yang penting untuk diperhatikan oleh negara demi membangun kualitas sumber daya manusianya. Dunia pendidikan harus terus berkembang untuk melakukan perbaikan yang berkelanjutan. Khususnya di Indonesia, perkembangan pendidikan terus dilakukan oleh pemerintah dengan berbagai cara guna mencerdaskan bangsanya. Selain itu, pendidikan juga menjadi salah satu rencana strategis pemerintah dalam rangka menghadapi MEA (Masyarakat Ekonomi ASEAN), sehingga Indonesia dapat bersaing dengan negara – negara lainnya.

Namun di tengah - tengah usaha perkembangan pendidikan tersebut, terdapat permasalahan yang masih dialami dunia pendidikan Indonesia. Masalah pemerataan dan akses pendidikan masih sulit dilaksanakan di Indonesia. Kondisi geografis menjadi salah satu faktor sulitnya pemecahan masalah tersebut, apalagi dengan kondisi geografis Indonesia yang merupakan negara kepulauan [1]. Hal tersebut menjadi tantangan utama bagi para pelaku dunia pendidikan untuk terus melakukan perkembangan demi kemajuan pendidikan Indonesia. Pemerataan kualitas pendidikan di setiap daerah merupakan langkah yang dapat dilakukan untuk peningkatan kualitas pendidikan di seluruh Indonesia [2]. Bukan hal yang buruk maupun sulit untuk dilakukan, tetapi hanya saja sistem pendidikan Indonesia belum dilaksanakan secara optimal [3].

Lembaga pendidikan merupakan suatu lembaga yang menyelenggarakan pendidikan dengan berusaha mencari langkah tepat untuk meningkatkan mutu pendidikan di sekolah dalam rangka memenuhi kebutuhan para siswa [4]. Banyaknya lembaga pendidikan yang hadir di Indonesia akan turut meramaikan dunia pendidikan serta menuntut setiap lembaga pendidikan memiliki strategi untuk menghadapi persaingan. Persaingan tersebut bersifat positif dalam mendorong kemajuan tiap lembaga pendidikan melakukan peningkatan kualitas pendidikan yang diselenggarakannya.

Penilaian sekolah dengan indikator tertentu dapat menganalisis kondisi mutu sekolah dalam suatu waktu. Menurut penelitian yang dilakukan Novita (2017), indikator penilaian sekolah dikatakan bermutu dari dinas pendidikan adalah memenuhi 8 SNP (Standar Nasional Pendidikan). Sedangkan, indikator mutu sekolah menurut orang tua siswa, yaitu akreditasi sekolah; kompetensi guru; lulusan siswa yang diterima sekolah/perguruan tinggi

unggul; prestasi siswa baik bidang akademik dan non akademik; hasil Ujian Nasional (UN); dan karakter siswa. Dari hasil penilaian sekolah tersebut dapat dijadikan sebagai dasar pemetaan mutu sekolah.

Yayasan pendidikan XYZ merupakan salah satu lembaga pendidikan Islam milik swasta di Indonesia yang melaksanakan kegiatan pendidikan formal, terdiri dari taman kanak – kanak, sekolah dasar, sekolah menengah pertama, sekolah menengah atas, hingga universitas. Dalam menghadapi persaingan dengan lembaga pendidikan lainnya serta mengikuti perkembangan zaman, sekolah perlu terus tumbuh dan berkembang. Salah satu rencana strategis yang dirancang oleh yayasan adalah melakukan program pemetaan sekolah–sekolah yang dimilikinya guna menganalisis kondisi mutu sekolah serta memetakan sekolah ke dalam tiga tingkatan, yaitu *Excellent School*; *Good School*; dan *Improvement School*. Tingkatan *Excellent School* akan berisi sekolah yang memiliki tingkatan kualitas paling baik. Tingkatan *Good School* akan berisi sekolah yang memiliki tingkatan kualitas yang cukup baik. Sedangkan, tingkatan *Improvement School* akan berisi sekolah yang memiliki tingkatan kualitas yang masih kurang baik. Hasil pemetaan tersebut dapat dijadikan yayasan sebagai dasar untuk evaluasi, sehingga dapat dirancang program yang sesuai masing-masing kebutuhan sekolah.

Berdasarkan permasalahan tersebut, dibutuhkan rancangan pemetaan untuk sekolah yang dimiliki yayasan pendidikan XYZ, sehingga dapat diketahui pengelompokan kualitas tiap sekolah yang dimilikinya. Pemetaan didasarkan dengan sejumlah indikator tertentu untuk penilaiannya. Dalam melakukan pemetaan dapat menggunakan peranan *data mining* berbasis metode algoritma *machine learning*. *Clustering* merupakan bagian dari *unsupervised learning* yang dapat dilakukan dengan metode *K-Means* dan *K-Medoids*. Dengan *clustering* tersebut, sekolah-sekolah akan memiliki karakteristik yang sama akan berkumpul pada *cluster* yang sama, sehingga dapat membentuk pemetaan sekolah. Selanjutnya, hasil *clustering* akan dijadikan sebagai *input* data pada metode *Decision Tree C4.5* sebagai *supervised learning* yang melakukan klasifikasi. Dengan klasifikasi tersebut, akan didapatkan pola pemetaan sekolah untuk memprediksi sekolah masuk ke dalam suatu *cluster* tertentu.

METODE

Desain, objek, dan waktu penelitian

Penelitian ini dilakukan pada Yayasan Pendidikan XYZ untuk memetakan serta melakukan simulasi prediksi sekolah pada jenjang SMA (Sekolah Menengah Atas). Waktu penelitian dilakukan pada Maret sampai Juli 2021.

Jenis pengumpulan data

Pengumpulan data dilakukan dengan menggunakan data primer dan *generate random number*. Data primer yang didapatkan dari hasil kuesioner dengan narasumber terkait, dalam hal ini terdapat tim pemetaan sekolah dari yayasan tersebut. Sedangkan, *generate random number* dilakukan menggunakan *Microsoft Excel* untuk asumsi 100 data sekolah dengan 9 indikator penilaian menggunakan *Generate random number* dilakukan karena dalam batasan waktu penelitian belum didapatkan data primer dari yayasan tersebut, sehingga digunakan alternatif penggunaan data *random number*. Dari proses pengumpulan data, dihasilkan profil yayasan dan data sekolah milik yayasan pendidikan XYZ.

Tahapan pengolahan dan analisis data

Pada tahapan pengolahan data, dimulai dari perancangan indikator pemetaan sekolah. Dalam melakukan perancangan, dilakukan FGD (*Focus Group Discussion*) hingga didapatkan validasi rancangan instrumen pemetaan sekolah dari narasumber terkait. Dari sejumlah indikator yang terdapat pada rancangan instrumen pemetaan dilakukan reduksi dengan metode *Value Engineering*. Reduksi dilakukan karena adanya keterbatasan waktu pengumpulan data maupun adanya data yang belum tersedia (belum didokumentasikan). Hasil reduksi indikator kemudian dilakukan pembobotan nilai dengan AHP (*Analytical Hierarchy Process*) untuk pemeringkatan indikator. Himpunan data selanjutnya dikalikan dengan bobot dan dilakukan normalisasi data.

Setelah itu, dilakukan pengklasteran dengan metode *K-Means* dan *K-Medoids Clustering*, serta membandingkan hasil *clustering* dari kedua metode tersebut dengan metode *Davies-Bouldin Index*. Kemudian dilanjutkan pengklasifikasian dengan metode *Decision Tree C4.5* untuk menemukan pola prediksi pemetaan sekolah dari hasil data *training* yang dihasilkan dari *clustering*. Perhitungan hasil akurasi pola *rules* hasil *decision tree* dilakukan dengan *Confusion Matrix*. Langkah terakhir, dilakukan simulasi pola *decision tree* untuk

memprediksi label *cluster* untuk data *testing* suatu sekolah

HASIL DAN PEMBAHASAN

Pengumpulan data

Dilakukan *generate random number* untuk asumsi data penilaian 100 sekolah berdasarkan 9 indikator yang ditetapkan dengan rubrik penilaian level 1 sampai 4. Himpunan data tersebut akan dijadikan sebagai input dalam pengolahan data penelitian ini. "IND 1" merupakan nilai dari indikator tingkat pemenuhan daya tampung; "IND 2" merupakan nilai dari tertib LPJ keuangan yang *outstanding*; "IND 3" merupakan nilai dari persentase murid diterima PTN; "IND 4" merupakan prestasi akademik lembaga Kemendikbud/resmi; "IND 5" merupakan akreditasi sekolah; "IND 6" merupakan rata-rata penilaian DP4 guru; "IND 7" merupakan prestasi non-akademik lembaga Kemendikbud/resmi; "IND 8" merupakan persentase pemenuhan sarana dan prasarana; dan "IND 9" merupakan rata-rata penilaian DP4 tenaga kependidikan. Berikut pada Tabel 1 merupakan himpunan data sekolah hasil *generate random number*.

Tabel 1. Himpunan Data

Sekolah	IND 1	IND 2	IND 3	IND 4	IND 5	IND 6	IND 7	IND 8	IND 9
1	1	3	3	2	4	2	3	3	4
2	4	3	3	4	2	4	4	4	4
3	3	4	2	1	4	1	2	1	1
4	2	2	3	3	3	3	3	1	3
5	2	2	1	2	1	3	2	2	2
6	3	1	1	4	2	3	4	1	3
7	3	1	4	4	4	4	4	3	4
8	1	3	2	4	2	2	1	4	4
9	4	4	3	2	1	4	1	4	2
10	3	2	3	3	4	2	4	3	3
...
100	2	4	2	1	2	3	2	3	2

Pengolahan data

Rancangan Instrumen Indikator Pemetaan Sekolah

Perancangan instrumen indikator pemetaan dilakukan untuk penilaian kualitas masing-masing sekolah. Instrumen tersebut merupakan alat penilaian yang terdiri dari indikator dan rubrik penilaian, terdapat empat bagian penilaian sekolah yang terdiri dari, "Sekolah Ideal"; "Sekolah Sehat"; "Sekolah Mutu"; dan "Sekolah Juara". Setiap bagian memiliki indikator utama yang berisikan sejumlah sub-indikator yang akan dinilai berdasarkan 4 level rubrik. Penilaian sekolah dikatakan "Sekolah Ideal" dilakukan dari indikator murid, SDM (Sumber Daya Manusia) guru, tenaga kependidikan, dan keuangan.

Indikator utama dalam penilaian “Sekolah Sehat” adalah fasilitas. Indikator utama dalam penilaian “Sekolah Mutu” adalah mutu. Sedangkan indikator utama dalam penilaian “Sekolah Juara” adalah prestasi murid, prestasi guru, dan prestasi sekolah.

Perancangan instrumen telah melalui beberapa kali tahapan revisi dari pendapat sekretariat yayasan, pejabat eselon, pengawas sekolah, sampai dengan para kepala sekolah. Setelah melalui tahapan revisi dilakukanlah validasi oleh narasumber terkait. Rancangan instrumen indikator pemetaan sekolah yang dihasilkan terdiri dari 34 sub-indikator.

Reduksi Indikator dengan Value Engineering

Rancangan instrumen indikator pemetaan sekolah direduksi dengan metode *Value Engineering*. Dengan adanya reduksi tersebut maka akan terpilih sub-indikator prioritas yang akan digunakan dalam pengolahan data selanjutnya. Pada penelitian ini, digunakan lima tahapan dalam metode *Value Engineering*, yaitu tahapan informasi, kreatif, analisa, pengembangan, dan penyajian [6].

Tahap informasi dilakukan dengan kegiatan studi literatur dan FGD (*Focus Group Discussion*). Tahap kreatif dilakukan dengan kegiatan perancangan instrumen pemetaan sekolah. Tahap analisa dilakukan dengan melakukan penilaian terhadap sub-indikator pada instrumen pemetaan sekolah melalui kuesioner yang diisi oleh narasumber terkait, serta dilakukan pengolahan terhadap penilaian tersebut sehingga diperoleh matrik hasil penilaian indikator. Tahap pengembangan dilakukan dengan membandingkan hasil penilaian sub-indikator, kemudian dilakukan pemilihan sub-indikator dengan nilai tertinggi, serta dilakukan

pertimbangan-pertimbangan tertentu melalui validasi pakar. Sedangkan tahap penyajian, dilakukan penyajian terhadap hasil sub-indikator yang terpilih.

Kuesioner yang disebarakan berbentuk penilaian tingkat kepentingan tiap *sub*-indikator pada instrumen pemetaan sekolah. Penilaian tingkat kepentingan tersebut dilakukan dengan skala semantik dari interval 1 (satu) sampai 7 (tujuh). Semakin sangat penting penilaian terhadap sub-indikator, maka bobot semakin besar yang mengarah ke kanan atau skala 7 (tujuh). Sebaliknya, semakin sangat tidak penting penilaian terhadap sub-indikator, maka bobot semakin kecil yang mengarah ke kiri atau skala 1 (satu).

Pemilihan indikator prioritas dilakukan berdasarkan nilai rata-rata skor yang tertinggi dari penilaian tingkat kepentingan oleh masing-masing pakar. Pakar yang melakukan penilaian kuesioner terdiri dari Pakar 1 yang merupakan Kepala Seksi Perencanaan dan Pengembangan Yayasan serta merupakan mantan Kepala SMA (Sekolah Menengah Atas); Pakar 2 yang merupakan Kepala Bidang Pengembangan Sumber Daya Manusia serta merupakan mantan Kepala SMA (Sekolah Menengah Atas); dan Pakar 3 yang merupakan salah satu anggota Bidang Pendidikan dan Pengembangan. Setelah dilakukan validasi pakar, didapatkan sembilan sub-indikator yang menjadi indikator prioritas, hasil tersebut telah mewakili masing-masing indikator utama. Berikut pada Tabel 2 merupakan rekapitulasi hasil indikator prioritas yang terpilih.

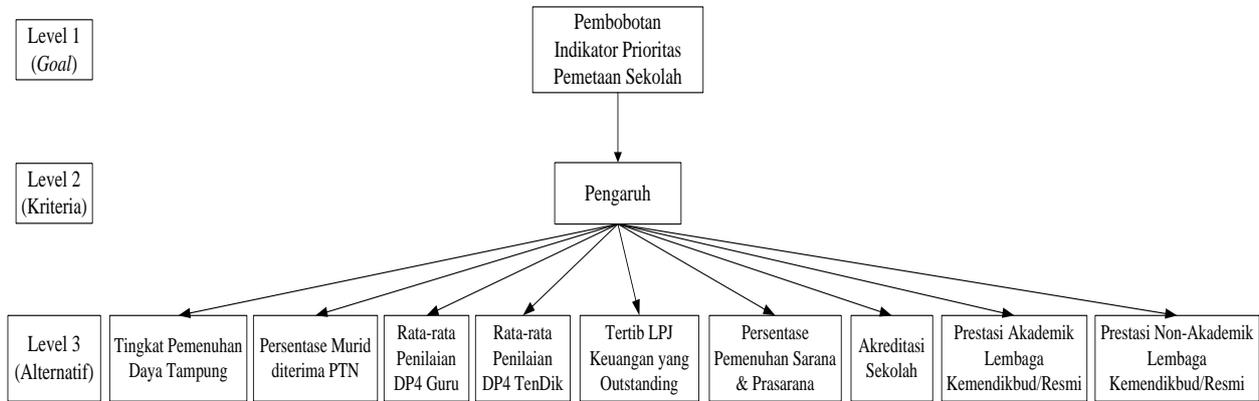
Tabel 2. Rekapitulasi Hasil Reduksi *Value Engineering*

Indikator Prioritas dalam Pemetaan Sekolah	Rubrik			
	1	2	3	4
Tingkat pemenuhan daya tampung sekolah	<79%	80-89%	90-99%	100%
Persentase jumlah murid yang diterima PTN (Perguruan Tinggi Negeri)	<10%	10-29%	30-49%	>= 50%
Rata-rata penilaian DP4 guru	0-49	50-64	65-79	80-100
Rata-rata penilaian DP4 tenaga kependidikan	0-49	50-64	65-79	80-100
Tertib Laporan pertanggungjawaban keuangan per tahun (yang belum dilaporkan- <i>Outstanding</i>)	> 3%	2,1 - 3%	1,1 - 2%	0-1%
Persentase pemenuhan standar sarana dan prasarana sekolah	0-25%	26-50%	51-75%	76-100%
Akreditasi sekolah	Belum Terakreditasi	C (Cukup)	B (Baik)	A (Unggul)
Prestasi bidang akademik lembaga penyelenggara Kemendikbud/Resmi	Juara Harapan	Perunggu/Juara 3	Perak/Juara 2	Emas/Juara 1
Prestasi bidang non-akademik lembaga penyelenggara Kemendikbud/Resmi	Juara Harapan	Perunggu/Juara 3	Perak/Juara 2	Emas/Juara 1

Pembobotan Indikator Prioritas dengan AHP (Analytical Hierarchy Process)

Pembobotan indikator prioritas dilakukan dengan metode AHP menggunakan bantuan *Software Super Decisions*. Pembobotan didasarkan pada pengaruh dari masing-masing indikator prioritas terhadap penilaian pemetaan sekolah. Dalam pembobotan

dengan metode AHP, dirancang struktur hirarki untuk mendefinisikan tujuan, kriteria, dan alternatif yang digunakan [7]. Berikut merupakan struktur hirarki dalam melakukan perhitungan dengan metode AHP.



Gambar 1. Struktur Hirarki

Penilaian matriks perbandingan berpasangan dilakukan oleh 2 pakar, yaitu Pakar 1 dan Pakar 2 yang sama pada metode *Value Engineering*. Sedangkan Pakar 3 tidak dapat melakukan pengisian kuesioner AHP ini dikarenakan keterbatasan waktu yang dimiliki pakar. Hasil penilaian masing-masing pakar telah konsisten dengan nilai *inconsistency* sebesar 0,077 untuk Pakar 1 dan 0,067 untuk Pakar 2. Selanjutnya penggabungan hasil penilaian kedua pakar dilakukan dengan perhitungan *geometric mean*.

tersebut menunjukkan bobot normalisasi dari tiap elemen indikator prioritas. Nilai *inconsistency* yang didapatkan dari *software* setelah penggabungan penilaian Pakar 1 dan Pakar 2 adalah 0.038 sehingga dapat dikatakan konsisten. Hasil pembobotan tiap indikator prioritas dari perhitungan metode AHP ini selanjutnya dikalikan dengan himpunan data yang telah didapatkan dari hasil pengumpulan data. Berikut pada Tabel 3 merupakan hasil himpunan data yang telah dikalikan dengan hasil bobot perhitungan AHP.

Gambar 2. Hasil Nilai Normalisasi

Tabel 3. Himpunan Data Hasil Pembobotan

Sekolah	IND 1	IND 2	IND 3	IND 4	IND 5	IND 6	IND 7	IND 8	IND 9
1	0.352	0.492	0.430	0.279	0.218	0.093	0.124	0.108	0.092
2	1.408	0.492	0.430	0.558	0.109	0.186	0.165	0.143	0.092
3	1.056	0.656	0.287	0.139	0.218	0.047	0.083	0.036	0.023
4	0.704	0.328	0.430	0.418	0.164	0.140	0.124	0.036	0.069
5	0.704	0.328	0.143	0.279	0.055	0.140	0.083	0.072	0.046
6	1.056	0.164	0.143	0.558	0.109	0.140	0.165	0.036	0.069
7	1.056	0.164	0.574	0.558	0.218	0.186	0.165	0.108	0.092
8	0.352	0.492	0.287	0.558	0.109	0.093	0.041	0.143	0.092
9	1.408	0.656	0.430	0.279	0.055	0.186	0.041	0.143	0.046
10	1.056	0.328	0.430	0.418	0.218	0.093	0.165	0.108	0.069
...
100	0.704	0.656	0.287	0.139	0.109	0.140	0.083	0.108	0.046

Normalisasi Data

Himpunan data yang telah di bobotkan dilakukan normalisasi untuk meratakan ukuran jarak dan keragaman data pada atribut yang akan digunakan pada perhitungan *clustering* sehingga menghindari data yang *outlier* [2]. Normalisasi data dilakukan dengan *Min-Max Normalization*. Berikut pada Tabel

Gambar 2 di atas menunjukkan hasil akhir metode AHP dengan *software Super Decisions*. Hasil

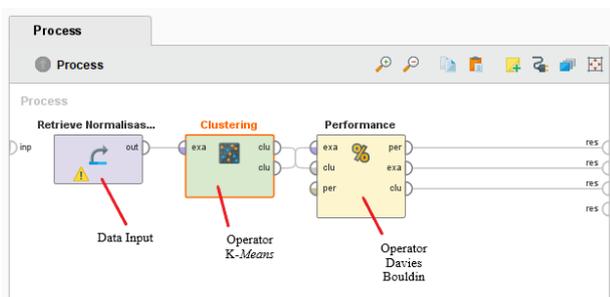
4 merupakan hasil himpunan data yang telah dilakukan normalisasi data.

Tabel 4. Himpunan Data Hasil Normalisasi

Sekolah	IND 1	IND 2	IND 3	IND 4	IND 5	IND 6	IND 7	IND 8	IND 9
1	0.238	0.339	0.294	0.185	0.141	0.051	0.073	0.061	0.050
2	1.000	0.339	0.294	0.386	0.062	0.118	0.103	0.087	0.050
3	0.746	0.457	0.191	0.084	0.141	0.017	0.043	0.009	0.000
4	0.492	0.220	0.294	0.285	0.102	0.084	0.073	0.009	0.033
5	0.492	0.220	0.087	0.185	0.023	0.084	0.043	0.035	0.017
6	0.746	0.102	0.087	0.386	0.062	0.084	0.103	0.009	0.033
7	0.746	0.102	0.398	0.386	0.141	0.118	0.103	0.061	0.050
8	0.238	0.339	0.191	0.386	0.062	0.051	0.013	0.087	0.050
9	1.000	0.457	0.294	0.185	0.023	0.118	0.013	0.087	0.017
10	0.746	0.220	0.294	0.285	0.141	0.051	0.103	0.061	0.033
...
100	0.492	0.457	0.191	0.084	0.062	0.084	0.043	0.061	0.017

Pengklasteran dengan K-Means

Hasil himpunan data yang telah dinormalisasi selanjutnya menjadi input untuk perhitungan K-Means dengan Software RapidMiner. Indikator 1 sampai indikator 9 pada himpunan data berperan sebagai atribut 1 sampai atribut 9 pada himpunan data. Berikut merupakan tampilan lembar process pada Software RapidMiner.



Gambar 3. Tampilan Lembar Proses K-Means

Tabel 5. Hasil K-Means Clustering

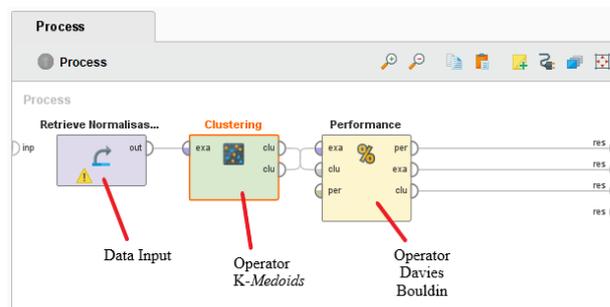
Sekolah	Cluster	Sekolah	Cluster
1	cluster_1	51	cluster_1
2	cluster_0	52	cluster_1
3	cluster_2	53	cluster_0
4	cluster_1	54	cluster_0
5	cluster_1	55	cluster_1
6	cluster_2	56	cluster_2
7	cluster_2	57	cluster_0
8	cluster_1	58	cluster_1
9	cluster_0	59	cluster_0
10	cluster_2	60	cluster_0
...
50	cluster_1	100	cluster_1

Pada Tabel 5 tersebut merupakan hasil K-Means clustering. Parameter yang digunakan pada Software RapidMiner, yaitu jumlah k = 3 karena cluster yang ingin dibentuk berjumlah 3, yaitu Excellent School; Good School; dan Improvement School. Kemudian, digunakan “numerical measures” dan dipilih

perhitungan “Euclidean Distance” untuk menghitung jarak antara data dengan pusat cluster. Cluster model yang terbentuk dari hasil K-Means clustering tersebut, yaitu cluster_0 berjumlah 23 items; cluster_1 berjumlah 45 items; dan cluster_2 berjumlah 32 items.

Pengklasteran dengan K-Medoids

Sama halnya dengan langkah dan parameter pada proses perhitungan K-Means, berikut pada Gambar 4 merupakan tampilan lembar process yang digunakan untuk perhitungan dengan K-Medoids Clustering. Perbedaan hasil clustering dari algoritma K-Means dan K-Medoids dapat terjadi karena perbedaan centroid. Centroid yang dihasilkan pada algoritma K-Means menggunakan nilai rata-rata tiap cluster [8], sedangkan algoritma K-Medoids menggunakan perwakilan objek yang representatif terhadap suatu cluster [9]. Perbedaan nilai data yang sangat kecil dapat menjadi salah satu akibat suatu data berpindah cluster saat menggunakan kedua algoritma yang berbeda tersebut.



Gambar 4. Tampilan Lembar Proses K-Medoids

Tabel 6. Hasil K-Medoids Clustering

Sekolah	Cluster	Sekolah	Cluster
1	cluster_2	51	cluster_0
2	cluster_1	52	cluster_0
3	cluster_2	53	cluster_1
4	cluster_0	54	cluster_1
5	cluster_2	55	cluster_0
6	cluster_1	56	cluster_1
7	cluster_1	57	cluster_1
8	cluster_0	58	cluster_2
9	cluster_1	59	cluster_1
10	cluster_1	60	cluster_1
...
50	cluster_2	100	cluster_2

Tabel 6 tersebut merupakan hasil perhitungan K-Medoids Clustering dari Software RapidMiner. Cluster model yang terbentuk dari hasil tersebut, yaitu cluster_0 berjumlah 20 items; cluster_1 berjumlah 44 items; dan cluster_2 berjumlah 36 items. Cluster model yang dihasilkan tersebut

berbeda dengan *cluster model* pada *K-Means clustering*, serta hasil *centroid* yang dihasilkan juga berbeda dengan *K-Means clustering*.

Perbandingan Hasil Pengklasteran dengan Davies Bouldin Index

Untuk menentukan hasil *clustering* yang paling baik antara algoritma *K-Means* dan *K-Medoids* digunakan perbandingan nilai *Davies Bouldin Index* (DBI). Perhitungan nilai DBI juga dilakukan dengan *Software RapidMiner* dengan menggunakan *operator Cluster Distance Performance* seperti yang ditunjukkan pada Gambar 3 dan Gambar 4. Berikut merupakan hasil nilai DBI dari masing-masing algoritma.



Gambar 5. Hasil DBI dari K-Means



Gambar 6. Hasil DBI dari K-Medoids

Gambar 5 dan Gambar 6 tersebut, penentuan *cluster* yang paling baik dari kedua metode tersebut dilakukan berdasarkan nilai DBI yang paling minimum [10], sehingga dapat dikatakan hasil *K-Medoids clustering* yang terpilih untuk usulan pemetaan sekolah. Nilai DBI yang semakin mendekati nilai nol akan semakin baik hasilnya, dalam artian semakin kecil nilai DBI tersebut semakin dekat kerapatan antar data pada *cluster* [11].

Selanjutnya, untuk mengetahui label dari tiap *cluster* yang dihasilkan, maka dilakukan analisis terhadap *centroid* atau pusat *cluster* yang dihasilkan dari masing-masing *cluster* [8], [12]. Berikut merupakan hasil *centroid* akhir yang dihasilkan pada *K-Medoids Clustering*.

Tabel 7. Hasil *Centroid* Akhir *K-Medoids*

	Cluster_0	Cluster_1	Cluster_2
IND 1	0.238	1.000	0.492
IND 2	0.102	0.220	0.457
IND 3	0.398	0.191	0.191
IND 4	0.386	0.386	0.084
IND 5	0.102	0.062	0.062
IND 6	0.051	0.051	0.084
IND 7	0.013	0.073	0.043
IND 8	0.035	0.061	0.061
IND 9	0.033	0.050	0.017
Rata-rata	1.357	2.094	1.491

Berdasarkan Tabel 7 tersebut, rata-rata *centroid* tertinggi adalah *cluster_1* yang diikuti *cluster_2* dan *cluster_0*. Dapat dikatakan sekolah yang memiliki nilai atribut indikator tertinggi akan berkumpul pada *centroid cluster* yang karakteristiknya sama-sama tinggi. Hasil tersebut dapat menunjukkan hasil *centroid* yang besar menunjukkan label tertinggi. Oleh karena itu, *cluster_1* memiliki label *Excellent School*; *cluster_2* memiliki label *Good School*; dan *cluster_0* memiliki label *Improvement School*. Berikut pada Tabel 8 merupakan hasil pemetaan sekolah dengan algoritma *K-Medoids*.

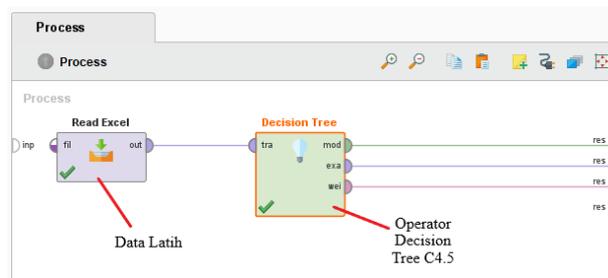
Tabel 8. Hasil Pemetaan Sekolah

Excellent School	Good School	Improvement School
Sekolah 2	Sekolah 1	Sekolah 4
Sekolah 6	Sekolah 3	Sekolah 8
Sekolah 7	Sekolah 5	Sekolah 11
Sekolah 9	Sekolah 13	Sekolah 16
Sekolah 10	Sekolah 14	Sekolah 23
Sekolah 12	Sekolah 17	Sekolah 27
Sekolah 15	Sekolah 18	Sekolah 34
Sekolah 22	Sekolah 19	Sekolah 36
Sekolah 24	Sekolah 20	Sekolah 43
Sekolah 25	Sekolah 21	Sekolah 44
Sekolah 29	Sekolah 26	Sekolah 47
Sekolah 30	Sekolah 28	Sekolah 51
Sekolah 31	Sekolah 32	Sekolah 52
Sekolah 33	Sekolah 37	Sekolah 55
Sekolah 35	Sekolah 38	Sekolah 64
Sekolah 39	Sekolah 42	Sekolah 67
Sekolah 40	Sekolah 48	Sekolah 84
Sekolah 41	Sekolah 50	Sekolah 94
Sekolah 45	Sekolah 58	Sekolah 96
Sekolah 46	Sekolah 61	Sekolah 97
Sekolah 49	Sekolah 68	
Sekolah 53	Sekolah 69	
Sekolah 54	Sekolah 72	
Sekolah 56	Sekolah 73	
Sekolah 57	Sekolah 74	
Sekolah 59	Sekolah 75	
Sekolah 60	Sekolah 76	
Sekolah 62	Sekolah 79	
Sekolah 63	Sekolah 82	
Sekolah 65	Sekolah 83	
Sekolah 66	Sekolah 87	
Sekolah 70	Sekolah 88	
Sekolah 71	Sekolah 89	
Sekolah 77	Sekolah 92	
Sekolah 78	Sekolah 99	
Sekolah 80	Sekolah 100	
Sekolah 81		
Sekolah 85		
Sekolah 86		
Sekolah 90		
Sekolah 91		
Sekolah 93		
Sekolah 95		
Sekolah 98		

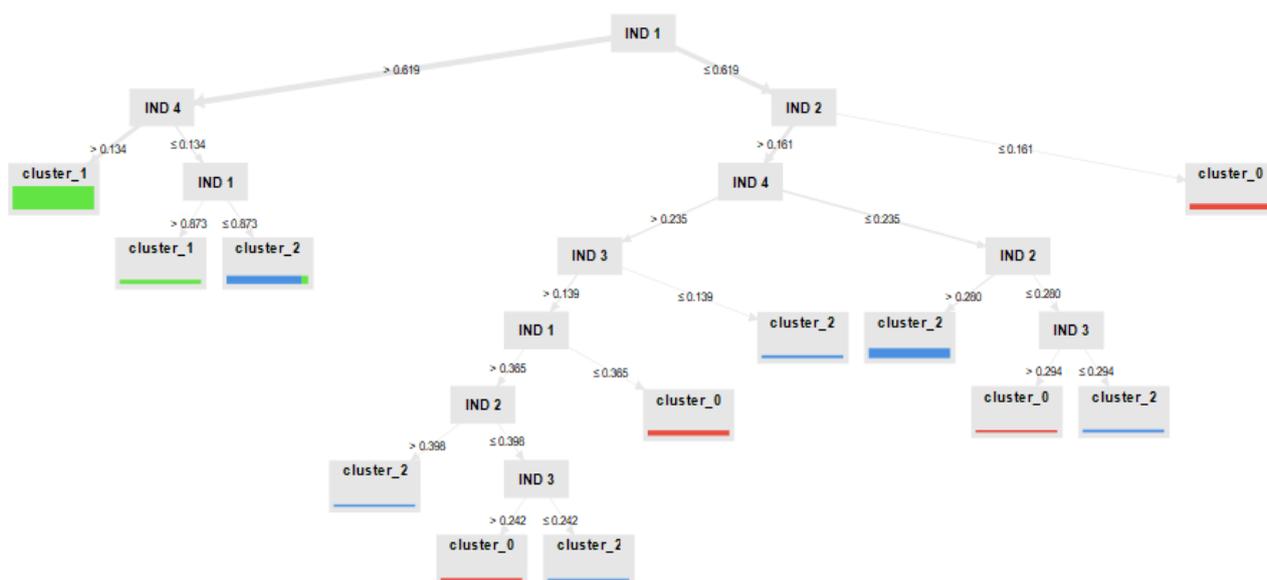
Pengklasifikasian dengan *Decision Tree C4.5*

Hasil *clustering* yang terpilih, yaitu *K-Medoids*, dijadikan sebagai data latih atau data *training* yang di input untuk menghitung algoritma *Decision Tree C4.5*. Berbeda dengan metode *K-Means* dan *K-Medoids* yang merupakan *unsupervised learning*, metode *Decision Tree C4.5* merupakan *supervised learning*. Data yang diinput pada metode *Decision Tree C4.5* berupa data yang telah memiliki label, dimana label tersebut merupakan hasil pengolahan *machine learning* sebelumnya [13], [14]. Pada penelitian ini, *cluster_0*, *cluster_1*, dan *cluster_2* merupakan label dari data latih hasil *clustering*. Perhitungan algoritma *Decision Tree C4.5*

dilakukan dengan menggunakan *Software RapidMiner* dengan lembar *process* yang digunakan sebagai berikut.



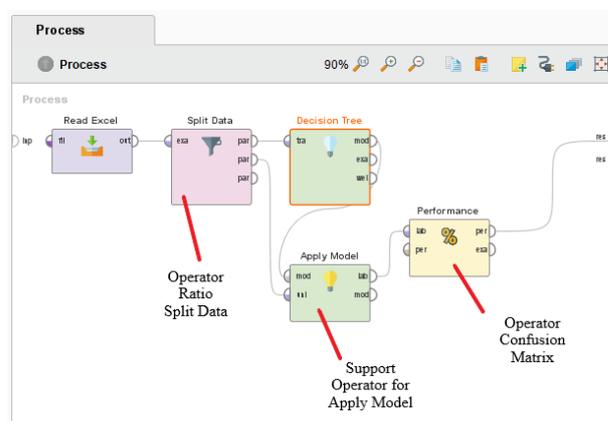
Gambar 7. Tampilan Lembar Proses *Decision Tree*



Gambar 8. Hasil *Decision Tree C4.5*

Gambar 8 tersebut menunjukkan bagan pohon keputusan dari hasil pengolahan *decision tree C4.5*. Hasil tersebut dapat dijadikan sebagai pola untuk melakukan prediksi klasifikasi suatu sekolah. Pohon keputusan yang dihasilkan hanya menampilkan *node* untuk IND 1 (indikator 1), IND 2 (indikator 2), IND 3 (indikator 3), dan IND 4 (indikator 4) disebabkan oleh hasil *attribute weights* keempat indikator tersebut mendominasi atribut lainnya, sehingga hanya dengan menganalisis keempat atribut indikator tersebut telah dapat mengklasifikasi suatu sekolah.

Kemudian, dilakukan pengujian akurasi hasil *decision tree* tersebut dengan menggunakan *Confusion Matrix* dengan *Software RapidMiner* seperti Gambar 9 berikut.



Gambar 9. Tampilan Lembar Proses *Confusion Matrix*

Untuk perhitungan *Confusion Matrix* dilakukan *split data* dengan rasio 50:50, 60:40, 70:30, 80:20, 90:10 antara data *training* dan data *testing* dari 100 data yang dimiliki. Berikut Tabel 9 merupakan hasil perhitungan akurasi *confusion matrix*.

Tabel 9. Hasil Akurasi *Confusion Matrix*

Rasio	Hasil Akurasi
50:50	84%
60:40	85%
70:30	90%
80:20	95%
90:10	90%

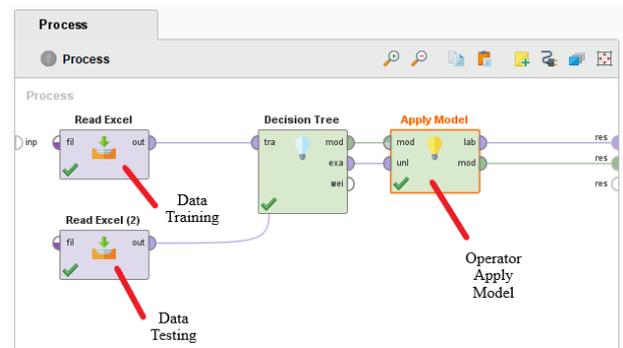
Berdasarkan Tabel 9 tersebut, perbandingan hasil akurasi terbesar berada pada rasio 80:20 sedangkan hasil akurasi terkecil berada pada rasio 50:50. Pada rasio 90:10 terdapat penurunan hasil akurasi dikarenakan terdapat data yang belum muncul pada proses pembelajaran sehingga pengklasifikasian tidak dapat dilakukan dengan baik oleh sistem [15]. Sedangkan, dari rasio 50:50 sampai dengan rasio 80:20 hasil akurasi cenderung mengalami peningkatan dikarenakan semakin banyak data *training* yang muncul pada sistem [15]. Semakin banyak data yang dikenali sistem pada keadaan tersebut mungkin *rules* data sudah muncul sebelumnya pada proses pembelajaran, sehingga sistem dapat mengklasifikasi dengan baik.

Selanjutnya, dilakukan simulasi dengan menggunakan model *decision tree* rasio 80:20 untuk data *testing* baru. Rasio 80:20 digunakan karena pada model tersebut memiliki hasil akurasi *confusion matrix* paling besar. Simulasi dilakukan guna memprediksi label *cluster* untuk data sekolah baru dengan menggunakan pola *decision tree* yang dihasilkan. Data *testing* yang digunakan dengan asumsi data sekolah ke-101 ditunjukkan sebagai berikut.

Tabel 10. Data *Testing* Baru

Sekolah	IND 1	IND 2	IND 3	IND 4	IND 5
101	0.746	0.457	0.191	0.386	0.141
	IND 6	IND 7	IND 8	IND 9	Cluster
	0.118	0.103	0.061	0.017	?

Untuk melakukan simulasi prediksi data *testing* baru dilakukan dengan *Software RapidMiner* dengan *process* yang ditunjukkan pada Gambar 10. Data *training* yang digunakan merupakan 80 data dari hasil *split data* pada model rasio 80:20. Sedangkan, data *testing* yang digunakan merupakan data *testing* baru untuk data sekolah ke-101.



Gambar 10. Tampilan Lembar Proses Simulasi

Dengan menggunakan proses pada Gambar 10, hasil simulasi prediksi ditunjukkan pada Gambar 11 di bawah ini. Hasil prediksi ditunjukkan pada kolom “prediction (Cluster)” yang disertai kolom “confidence(cluster_1)” sebesar 1. Hasil prediksi tersebut menunjukkan bahwa data *testing* baru untuk asumsi sekolah ke-101 memiliki label *cluster_1*. Hasil “confidence(cluster_1)” sebesar 1 menunjukkan bahwa data *testing* baru tersebut masuk ke *cluster_1*.

Row No.	Cluster	prediction(Cluster)	confidence(cluster_2)	confidence(cluster_1)	confidence(cluster_0)	IND 1	IND 2
1	?	cluster_1	0	1	0	0.746	0.457

Gambar 11. Hasil Simulasi Prediksi

KESIMPULAN

Berdasarkan pengumpulan dan pengolahan data yang dilakukan pada penelitian ini, dapat ditarik kesimpulan sebagai berikut:

Instrumen indikator pemetaan sekolah dirancang untuk melakukan penilaian terhadap kualitas masing-masing sekolah yang dimiliki yayasan

pendidikan XYZ. Instrumen tersebut merupakan alat penilaian yang terdiri dari indikator dan rubrik yang terdiri dari bagian Sekolah Ideal, Sekolah Sehat, Sekolah Mutu, dan Sekolah Juara.

Hasil reduksi dengan *Value Engineering* didapatkan indikator prioritas untuk atribut himpunan data, yaitu tingkat pemenuhan daya tampung; tertib LPJ keuangan yang *outstanding*; persentase murid

diterima PTN; prestasi akademik lembaga Kemendikbud/resmi; akreditasi sekolah; rata-rata penilaian DP4 guru; prestasi non-akademik lembaga Kemendikbud/resmi; persentase pemenuhan sarana dan prasarana; dan rata-rata penilaian DP4 tenaga kependidikan. Kemudian, dengan menggunakan metode AHP (*Analytical Hierarchy Process*) didapatkan bobot atribut tersebut berturut-turut, yaitu 0.35190; 0.16408; 0.14338; 0.13942; 0.05453; 0.04654; 0.04137; 0.03584; dan 0.02295.

Hasil simulasi rancangan pemetaan yayasan pendidikan XYZ dengan *K-Means Clustering* didapatkan pada cluster_0 berjumlah 23 items; cluster_1 berjumlah 45 items; dan cluster_2 berjumlah 32 items. Sedangkan, dengan *K-Medoids Clustering* didapatkan hasil cluster_0 berjumlah 20 items; cluster_1 berjumlah 44 items; dan cluster_2 berjumlah 36 items.

Hasil perbandingan *clustering* menggunakan metode Davies Bouldin Index didapatkan nilai DBI dari algoritma *K-Means* adalah sebesar 0.175. Sedangkan, nilai DBI dari algoritma *K-Medoids* adalah sebesar 0.112 sehingga hasil *K-Medoids* yang terpilih untuk usulan pemetaan sekolah.

Pola prediksi pemetaan sekolah dengan metode *Decision Tree C4.5* dihasilkan melalui bagan pohon keputusan. Sedangkan, hasil simulasi prediksi dengan menggunakan model rasio 80:20 didapatkan bahwa data *testing* baru dengan asumsi sekolah ke-101 masuk ke cluster_1 dengan tingkat akurasi sebesar 95%.

UCAPAN TERIMA KASIH

Terima kasih kepada Bapak Budi Aribowo selaku pembimbing pada penelitian ini, terima kasih kepada Yayasan XYZ khususnya Tim Pemetaan Sekolah atas kesediaan menjadi narasumber sekaligus objek penelitian serta terimakasih kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Al Azhar Indonesia melalui pendanaan Grant Internal Tahun Anggaran 2021.

REFERENSI

- [1] D. N. L. Laksana, "Pendidikan Berkualitas dan Berkelanjutan di Era Pembelajaran Abad 21," *J. Ilm. Pendidik. Citra Bakti*, vol. 5, no. 1, pp. 1–5, 2018.
- [2] G. S. Nugraha, Hairani, and R. F. P. Ardi, "Aplikasi Pemetaan Kualitas Pendidikan di

- Indonesia Menggunakan Metode K-Means," *J. Matrik*, vol. 17, no. 2, pp. 13–23, 2018.
- [3] N. Afifah, "Sistem Pendidikan di Indonesia," 2020.
- [4] D. M. D. Kamayuda, "Perencanaan Strategi Bersaing Sekolah dalam Meningkatkan Jumlah Peserta Didik Baru di Salah Satu Sekolah Swasta Salatiga," *J. Manaj. Pendidik.*, vol. 3, no. 1, 2016.
- [5] L. Novita, "Indikator Mutu Sekolah Menurut Perspektif Orang Tua Siswa di SMP Negeri 2 Bantul," *J. Kebijak. Pendidik.*, vol. VI, no. 2, pp. 184–193, 2017.
- [6] M. F. Pasaribu and R. Puspita, "Tahap Informasi, Kreatif], dan Analisa Pada Rekayasa Nilai Untuk Meningkatkan Kualitas Pelayanan Hotel," *Ind. Eng. J.*, vol. 5, no. 2, pp. 46–51, 2016.
- [7] A. E. Munthafa and H. Mubarak, "Penerapan Metode Analytcal Hierarchy Process dalam Sistem Pendukung Keputusan Penentuan Mahasiswa Berprestasi," *J. Siliwangi*, vol. 3, no. 2, pp. 192–201, 2017.
- [8] A. K. Wardhani, "Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Kajen Pekalongan," *J. Transform.*, vol. 14, no. 1, pp. 30–37, 2016.
- [9] F. Farahdinna, I. Nurdiansyah, A. Suryani, and A. Wibowo, "Perbandingan Algoritma K-means dan K-medoids dalam Klasterisasi Produk Asuransi Perusahaan Nasional," *J. Ilm. FIFO*, vol. XI, no. 2, pp. 208–214, 2019.
- [10] A. Badruttamam and D. A. I. Maruddani, "Penerapan Analisis Klaster K -Modes Dengan Validasi Davies Bouldin Index Dalam Menentukan Karakteristik Kanal Youtube Di Indonesia (Studi Kasus: 250 Kanal YouTube Indonesia Teratas Menurut Socialblade)," vol. 9, pp. 263–272, 2020.
- [11] A. F. Muhammad, "Klasterisasi Proses Seleksi Pemain Menggunakan Algoritma K-Means," pp. 1–5, 2015.
- [12] A. P. Windarto, U. Indriani, M. R. Raharjo, and L. S. Dewi, "Bagian 1 : Kombinasi Metode Klastering dan Klasifikasi (Kasus Pandemi Covid-19 di Indonesia)," vol. 4, pp. 855–862, 2020.
- [13] Galih, "Data Mining di Bidang Pendidikan untuk Analisa Prediksi Kinerja Mahasiswa dengan Komparasi 2 Model Klasifikasi pada STMIK Jabar," *J. Teknol. Sist. Inf. dan Apl.*, vol. 2, no. 1, pp. 23–30, 2019.
- [14] P. Kasih, "Pemodelan *Data Mining Decision Tree* Dengan *Classification Error* Untuk

- Seleksi Calon Anggota Tim Paduan Suara,” *Innov. Res. Informatics*, vol. 1, no. 2, pp. 63–69, 2019.
- [15] A. A. Mahardhika, R. Saptono, and R. Anggrainingsih, “Sistem Klasifikasi Feedback Pelanggan dan Rekomendasi Solusi Atas Keluhan di UPT PUSKOM UNS dengan Algoritma *Naive Bayes Classifier* dan *Cosine Similarity*,” *J. ITSMART*, vol. 4, no. 1, pp. 36–42, 2015.