

Klasterisasi Hasil Ujian Nasional SMA/MA dengan Algoritma K-Means

IWA. Suputra^{1,*}, IM. Candiasa², IPP. Suryawan³

¹²³*Program Studi S1 Pendidikan Matematika FMIPA, Universitas Pendidikan Ganesha Singaraja, Indonesia*

*Corresponding author: arlansuputra@gmail.com

Abstrak

Penelitian bertujuan untuk mengetahui bagaimana klasterisasi kualitas pendidikan SMA/MA berdasarkan hasil ujian nasional SMA/MA provinsi di Indonesia tahun ajaran 2018/2019 dengan algoritma K-Means. Algoritma K-Means adalah metode klasterisasi non hirarki yang mempartisi objek ke dalam beberapa kelompok (klaster). Teknik ini bertujuan untuk meminimumkan variasi di dalam suatu klaster dan memaksimumkan variasi antar klaster. Penelitian ini adalah penelitian *data mining* (penambangan data). Data penelitian diperoleh dari *website* resmi Pusat Penilaian Pendidikan dan Kebudayaan Kementerian Pendidikan dan Kebudayaan Republik Indonesia. Data yang diperoleh berisi rerata variabel yang diujikan dalam ujian nasional. Prosedur penambangan data yang ditempuh dalam penelitian ini yaitu pembersihan data, integrasi data, pemilihan data, transformasi data, penambangan data, evaluasi pola dan presentasi pengetahuan. Hasil klasterisasi yang diperoleh dalam penelitian ini adalah terdapat 2 provinsi yang dikategorikan ke dalam klaster 1, 10 provinsi dikategorikan ke dalam klaster 2, 11 provinsi dikategorikan ke dalam klaster 3 dan 12 provinsi dikategorikan ke dalam klaster 4. Hasil evaluasi dari algoritma K-Means menghasilkan nilai evaluasi Partition Coefficiens Index (PCI) 0,81.

Kata-kata kunci : *klasterisasi, kualitas, ujian nasional, Algoritma K-Means*

Abstract

This study aimed to find out how to cluster the quality of SMA / MA education based on the provincial SMA / MA national exam results in Indonesia in the 2018/2019 school year with an algorithm K-Means. The K-Means algorithm is a non-hierarchical clustering method that divides objects into groups (cluster). This method aims to minimize variation within a group and maximize variation between groups. This research is a data mining study. Data study is obtained from the official website of the Education and Culture Assessment Center of the Ministry of Education and Culture of the Republic of Indonesia. The data obtained contains the average of variables tested in the national exam. Data mining procedures adopted in this study are data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation. Clustering results obtained in this study there are 2 provinces which are categorized into cluster 1, 10 provinces are categorized into cluster 2, 11 are categorized into cluster 3 and 12 provinces are categorized into cluster 4. The evaluation resulted from the K-Means algorithm produced an evaluation value of Partition Coefficiens Index (PCI) of 0.81.

Keywords : *clustering, quality, national exams, K-Means Algorithm*

PENDAHULUAN

Ujian Nasional (UN) merupakan salah satu bentuk evaluasi yang dilakukan oleh pemerintah Indonesia untuk mengetahui kemampuan kognitif siswa secara nasional pada

jenjang sekolah dasar hingga sekolah menengah atas dan sederajat. Ujian nasional yang telah dilaksanakan beberapa tahun terakhir bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu. Hasilnya digunakan sebagai: (1) Pemetaan mutu satuan dan/atau program pendidikan; (2) seleksi masuk jenjang pendidikan berikutnya; dan (3) pembinaan dan pemberian bantuan kepada satuan pendidikan dalam upaya peningkatan mutu pendidikan.

Hasil studi *Programme for International Student Assessment* (PISA) tahun 2018 yang membandingkan kemampuan matematika, membaca dan kinerja sains setiap anak di 79 negara. Untuk kategori matematika, Indonesia berada di peringkat 73, untuk kemampuan membaca Indonesia berada di peringkat 74, lalu untuk kategori kinerja sains, Indonesia berada di peringkat 71. Berdasarkan laporan tersebut, performa Indonesia terlihat menurun jika dibandingkan dengan laporan PISA 2015 (Tohir, 2019). Selain dalam hasil studi internasional, kemampuan siswa, khususnya siswa SMA/MA juga terlihat pada rerata ujian nasional beberapa tahun terakhir. Dilansir dari Pusat Penilaian Pendidikan Kementerian Pendidikan dan Kebudayaan, rerata skor UN SMA/MA mengalami penurunan dari tahun 2016 sampai tahun 2018, kemudian mengalami peningkatan di tahun 2019. Walaupun demikian rerata yang diperoleh setiap tahun masih dikategorikan kecil. Peningkatan dan penurunan rerata nilai menunjukkan bahwa mutu pendidikan di Indonesia, khususnya pada jenjang SMA/MA masih belum stabil dan perlu ditingkatkan.

Berdasarkan paparan di atas, setidaknya memberikan alasan yang kuat bagi pemerintah sebagai penanggung jawab penyelenggara pendidikan untuk melakukan evaluasi dan segera berbenah terhadap proses pendidikan yang dijalankan. Menanggapi hal tersebut, harus dianalisis mengenai faktor-faktor yang dapat dijadikan pedoman atau pertimbangan dalam peningkatan kualitas dan pemerataan pendidikan di Indonesia. Melalui teknik analisis tertentu, kita dapat menggali informasi yang bermanfaat dan dapat memberikan informasi untuk khalayak luas maupun pemerintah. Untuk dapat menggali informasi dari data suatu data, diperlukan suatu proses pengolahan data. Salah satu teknik olah data yang dapat digunakan adalah teknik *data mining* (penambangan data).

Penambangan data menganalisis kumpulan data hasil observasi, menemukan hubungan antar variabel dan dapat merangkum data dengan cara yang baru sehingga dapat berguna dan dimengerti bagi pengguna (R.Mythily dkk, 2015). Klasterisasi adalah salah satu metode data mining yang mengelompokkan data atau objek ke dalam klaster sehingga dalam setiap klaster akan berisi data yang mirip atau memiliki karakteristik yang sama (Irwan, 2012). Secara sekilas, mirip dengan klasifikasi, namun klasterisasi belum didefinisikan sebelum dijalankan prosedur penambangan data, artinya pada klasterisasi tidak diketahui (tidak terdapat informasi) bagaimana data harus dikelompokkan, sehingga memerlukan prosedur tertentu (algoritma) untuk menghasilkan suatu kelompok sedangkan pada klasifikasi sudah terdapat informasi mengenai bagaimana data tersebut dikelompokkan, kemudian dilakukan *training* pada sistem dengan data yang sudah diberikan label (ke dalam kelompok manakah kelompok tersebut dikelompokkan), selanjutnya sistem akan mengklasifikasikan data-data yang baru ke dalam kelompok yang ada. Salah satu metode klasterisasi adalah Algoritma K-Means. Algoritma K-Means adalah metode klasterisasi yang dapat mempartisi data ke dalam beberapa kelompok (klaster) (Khomarudin, 2016).

Data hasil ujian nasional masing-masing provinsi di Indonesia dapat dikatakan termasuk *database* yang cukup besar. Data yang terdapat di dalamnya pun adalah data numerik yang dikategorikan bervariasi karena terdapat beberapa variabel (mata pelajaran)

yang dinilai serta kemampuan peserta didik di masing-masing daerah tentunya berbeda-beda. Algoritma K-Means sangat tepat digunakan pada data ujian nasional karena algoritma K-Means selain merupakan algoritma klasik yang sederhana dan cepat untuk menyelesaikan masalah kluster terutama untuk data numerik juga sangat fleksibel dan efisien untuk ukuran data yang cukup besar dan menyebar (bervariasi) (Yadav & Sharma, 2012). Selain itu, dengan algoritma K-Means, kita bisa menentukan jumlah kluster yang diinginkan, sehingga nantinya bisa ditentukan provinsi mana yang termasuk ke dalam kluster dengan kategori sangat baik, baik, cukup, maupun kurang. Dari analisa data hasil UN, bisa dilakukan evaluasi serta mengambil pertimbangan dan kebijakan untuk mengatasi masalah-masalah pendidikan, dalam hal ini berkaitan dengan pemetaan mutu pendidikan.

Beberapa penelitian telah membuktikan bahwa algoritma K-Means efektif diterapkan dalam pengelompokan data. Penelitian-penelitian tersebut adalah penelitian oleh Agil Aditya, Ivan Jovian, dan Betha Nurina Sari (2020) diperoleh bahwa algoritma K-Means cukup efektif diterapkan dalam pengelompokan, khususnya pada data ujian nasional. Adapun Penelitian oleh Tria Pratiwi Sutriyani, Amril Mutoi Siregar, dan Dwi Sulistya Kusumaningrum (2018) diperoleh algoritma K-Means dapat diterapkan pada data ujian nasional dengan menerapkan prosedur penambangan data dengan baik. Kemudian, penelitian Ninik Tri Hartanti (2018) diperoleh bahwa algoritma K-Means secara efektif dapat diterapkan pada data ujian nasional untuk menentukan kelompok belajar siswa. Penelitian dengan menggunakan algoritma K-Means juga dilakukan oleh oleh Evi Dewi Sri Mulyani, Susanto, Yoga Handoko Agustin, dan Nensi Mardhiani Surgawi (2018) diperoleh bahwa algoritma K-Means cukup efektif diterapkan dalam pengelompokan dengan nilai k (banyak kluster) ditentukan sesuai dengan kebutuhan.

Berdasarkan uraian di atas, diperoleh bahwa algoritma K-Means telah banyak digunakan dan efektif untuk diterapkan dalam masalah pengelompokan. Banyak penelitian tentang klusterisasi khususnya di bidang pendidikan yang telah dilakukan seperti penelitian-penelitian di atas, namun belum ada bukti empiris tentang klusterisasi kualitas pendidikan SMA/MA berdasarkan capaian nilai ujian nasional SMA/MA. Oleh karena itu, penulis tertarik untuk melakukan penelitian untuk klusterisasi kualitas pendidikan SMA/MA berdasarkan hasil UN masing-masing provinsi di Indonesia. Mengingat keterbatasan waktu dan tenaga, peneliti hanya menggunakan data ujian nasional tahun 2018/2019. Melalui penelitian ini, diharapkan bisa memberikan kontribusi ilmiah khususnya dalam penerapan algoritma K-Means.

METODE

Penelitian ini merupakan penelitian penambangan data. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari *website* resmi Pusat Penilaian Pendidikan Kementerian Pendidikan dan Kebudayaan Republik Indonesia berupa hasil capaian nilai ujian nasional SMA/MA tahun pelajaran 2018/2019 provinsi di Indonesia. Prosedur penambangan data yang digunakan dalam penelitian ini adalah pembersihan data, integrasi data, pemilihan data, transformasi data, proses *data mining*, evaluasi pola dan presentasi pengetahuan (Khomarudin, 2016). Metode *data mining* yang digunakan dalam penelitian ini adalah klusterisasi dengan Algoritma K-Means. Algoritma K-Means merupakan metode klusterisasi non hirarki yang mempartisi objek ke dalam beberapa kelompok (kluster). Teknik ini bertujuan untuk meminimumkan variasi di dalam suatu kluster dan memaksimumkan variasi antar kluster. Adapun tahapan algoritma K-Means yang ditempuh

yaitu: (1) Menentukan jumlah kluster (k); (2) Memilih secara acak titik k untuk dijadikan pusat kluster; (3) Menghitung jarak data yang lain dengan pusat kluster dengan perhitungan jarak *euclidean*; (4) Mengelompokkan data ke dalam kluster dengan jarak *euclidean* minimal; (5) Menghitung kembali pusat kluster dengan menghitung nilai rata-rata data pada setiap kluster; (6) Menghitung kembali jarak antar data dengan pusat kluster yang baru seperti pada langkah 3; dan (7) Mengulangi langkah hingga tidak ada perubahan pada pusat kluster, (Prasetyo, 2012).

HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengetahui bagaimana klusterisasi kualitas pendidikan SMA/MA berdasarkan hasil ujian nasional SMA/MA provinsi di Indonesia tahun ajaran 2018/2019 dengan algoritma K-Means. Variabel yang digunakan sebagai indikator klusterisasi adalah nilai bahasa Indonesia, bahasa Inggris, matematika, fisika, kimia, dan biologi. Dari rerata nilai masing-masing mata pelajaran yang diperoleh masing-masing provinsi, tiap provinsi akan dikelompokkan menjadi empat kluster dengan kategori pencapaian hasil ujian nasional sangat baik, baik, cukup, dan kurang. Pada tahap pembersihan data (*data cleaning*) tidak ditemukan data yang tidak konsisten atau data yang tidak relevan, oleh karena itu prosedur selanjutnya dapat dilakukan. Kemudian tahap Integrasi data (*data integration*), penelitian ini menggunakan data yang diperoleh dari *website* resmi Pusat Penilaian Pendidikan Kemendikbud. Di dalam *website* ini sudah terdapat rerata hasil UN masing-masing provinsi di Indonesia untuk setiap mata pelajaran yang diujikan, sehingga tidak perlu dilakukan penggabungan data dari berbagai sumber. Kemudian tahap pemilihan data (*data selection*) adalah memilih data yang diperlukan dalam proses analisis. Data ini memiliki 34 data yang terdiri dari 10 atribut. Keterangan data dapat dilihat pada tabel 1 berikut.

Tabel 1. Dataset Awal

Atribut	Tipe Data	Keterangan
Kode	Numerik	Kode Provinsi
Nama Provinsi	Nominal	Nama Provinsi
Jumlah Satuan Pendidikan	Numerik	Banyak satuan pendidikan di provinsi
Jumlah Peserta	Numerik	Banyak peserta ujian
Rerata nilai Bahasa Indonesia	Numerik	Hasil rerata nilai Bahasa Indonesia tiap provinsi
Rerata nilai Bahasa Inggris	Numerik	Hasil rerata nilai Bahasa Inggris tiap provinsi
Rerata nilai Matematika	Numerik	Hasil rerata nilai Matematika tiap provinsi
Rerata nilai Fisika	Numerik	Hasil rerata nilai Fisika tiap provinsi
Rerata nilai Kimia	Numerik	Hasil rerata nilai Kimia tiap provinsi
Rerata nilai Biologi	Numerik	Hasil rerata nilai Biologi tiap provinsi

Dari *dataset* awal, dilakukan proses seleksi, dimana proses seleksi bertujuan untuk memilih atribut-atribut yang diperlukan dalam proses analisis. Atribut atribut yang diperlukan antara lain nama provinsi, rerata nilai bahasa Indonesia, rerata nilai bahasa Inggris, rerata nilai matematika, rerata nilai fisika, rerata nilai kimia, dan rerata nilai biologi. Hasil seleksi dapat dilihat pada tabel 2 berikut.

Tabel 2. Dataset Hasil Seleksi

Atribut	Tipe Data	Keterangan
Nama Provinsi	Nominal	Nama Provinsi
Rerata nilai Bahasa Indonesia	Numerik	Hasil rerata nilai Bahasa Indonesia tiap provinsi
Rerata nilai Bahasa Inggris	Numerik	Hasil rerata nilai Bahasa Inggris tiap provinsi
Rerata nilai Matematika	Numerik	Hasil rerata nilai Matematika tiap provinsi
Rerata nilai Fisika	Numerik	Hasil rerata nilai Fisika tiap provinsi
Rerata nilai Kimia	Numerik	Hasil rerata nilai Kimia tiap provinsi
Rerata nilai Biologi	Numerik	Hasil rerata nilai Biologi tiap provinsi

Berikut adalah data yang akan digunakan pada tahap data mining.

Tabel 3. Capaian Nilai Ujian Nasional SMA/MA Tahun Pelajaran 2018/2019

Nama Provinsi	Rerata Nilai Pada Mata Uji					
	Bahasa Indonesia	Bahasa Inggris	Matematika	Fisika	Kimia	Biologi
DKI Jakarta	80,17	73,97	52,45	60,3	61,83	61,07
Jawa Barat	70,88	55,23	38,65	44,74	50,32	50,24
Jawa Tengah	77,91	58,65	44,65	51,76	56,54	57,03
DI Yogyakarta	82,06	67,97	50,86	55,37	60,27	62,6
Jawa Timur	73,26	56,54	41,92	48,03	53,89	54,52
Aceh	56,25	41,72	32,36	37,66	41,69	42,34
Sumatera Utara	61,85	47,85	36,4	41,62	47,27	46,02
Sumatera Barat	72,85	52,82	41,1	47,24	52,57	52,3
Riau	68,71	50,47	37,24	42,83	48,27	49,04
Jambi	66,22	48,27	36,05	41,88	46,71	47,72
Sumatera Selatan	63,75	46,92	35,33	40,56	46,11	46,87
Lampung	68,78	48,58	36,18	42	47,23	48,6
Kalimantan Barat	70,06	50,59	36,54	43,41	48,78	48,94
Kalimantan Tengah	67,38	48,07	33,88	39,24	44,53	47,1
Kalimantan Selatan	74,47	52,17	37,32	43,54	48,86	52,36
Kalimantan Timur	74,24	55,83	38,26	46,07	47,87	51,8
Sulawesi Utara	62,49	49,39	33,03	38,7	40,7	44,76
Sulawesi Tengah	62,96	43,86	32,54	38,79	43,23	45,26
Sulawesi Selatan	62,34	44,54	33,88	39,34	43,86	44,82
Sulawesi Tenggara	63,91	44,92	34,48	42,92	45,87	46,43
Maluku	55,87	42,25	33,22	38,97	39,94	45,83
Bali	74,85	59,55	41,08	47,87	55,78	55,67
Nusa Tenggara Barat	62,96	44,83	33,22	40,94	44,14	44,39
Nusa Tenggara Timur	62,05	42,51	31,68	37,63	40,79	44,46
Papua	57,64	44,6	33,23	40,49	41,13	43,4
Bengkulu	68,09	47,93	37,05	42,15	46,84	49,57

Maluku Utara	52,12	41,98	33,52	38,99	41,41	45,66
Bangka Belitung	74,37	53,4	39,9	46,59	49,35	51,72
Gorontalo	64,55	44,66	33,31	40,77	42,54	45,82
Banten	68,9	55,49	39,28	47,55	50,73	49,47
Kepulauan Riau	74,41	60,53	42,25	48,44	54,4	52,69
Sulawesi Barat	61,85	41,61	31,61	38,2	40,5	43,16
Papua Barat	61,24	46,83	37,49	43,9	48,24	47,59
Kalimantan Utara	70,31	49,64	34,3	38,6	44,33	48,74

Pada tahap Transformasi data dilakukan perubahan nama variabel & normalisasi data untuk memudahkan proses data mining seperti yang terlihat pada tabel 3 berikut.

Tabel 3. Pergantian Nama Variabel Data

Nama variabel lama	Nama variabel baru
Nama Provinsi	X
Rerata nilai Bahasa Indonesia	X ₁
Rerata nilai Bahasa Inggris	X ₂
Rerata nilai Matematatika	X ₃
Rerata nilai Fisika	X ₄
Rerata nilai Kimia	X ₅
Rerata nilai Biologi	X ₆

Dalam penelitian ini tidak dilakukan normalisasi data karena X₁, X₂, X₃, X₄, X₅, dan X₆ adalah nilai rerata UN yang berada pada rentang 0 – 100, sehingga tidak akan ada parameter yang mendominasi dalam perhitungan jarak antar data yang akan dilakukan ada tahap selanjutnya. Setelah tranformasi data dilakukan kemudian dilanjutkan teknik penambangan (*data mining*), dimana teknik yang akan digunakan adalah teknik klasterisasi dengan menggunakan algoritma K-Means. Pertama tentukan jumlah kluster. Jumlah kluster yang digunakan adalah 4 ($k = 4$). Pemilihan jumlah kluster bisa menggunakan aturan *thum* dengan formula $k \approx \sqrt{n/2}$ (Kordinariya & R. Makwan, 2013). Kluster 1 menunjukkan pencapaian hasil UN dikategorikan sangat baik, kluster 2 menunjukkan pencapaian hasil UN baik, kluster 3 pencapaian hasil UN cukup dan kluster 4 menunjukkan pencapaian hasil UN kurang. Langkah kedua adalah menentukan titik pusat kluster secara acak, kemudian hitung jarak setiap data dengan pusat kluster dengan jarak *euclidean* sampai data dikelompokan sesuai dengan jarak *euclidean* minimal. Lakukan beberapa kali iterasi sampai titik pusat dari setiap kluster tidak berubah dan tidak ada lagi data yang berpindah dari satu kluster ke kluster yang lain. Dalam penelitian ini, titik pusat kluster tidak mengalami perubahan pada iterasi ketiga. Berikut titik pusat kluster yang dihasilkan, disajikan Tabel 4.

Tabel 4. Titik Pusat Klaster pada Iterasi Terakhir

Klaster	Kualitas	X					
		X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
Klaster 1	Sangat baik	81,115	70,97	51,655	57,835	61,05	61,835
Klaster 2	Baik	73,614	56,021	40,441	47,183	52,031	52,78
Klaster 3	Cukup	66,390	48,188	35,903	41,737	46,743	47,874
Klaster 4	Kurang	60,098	43,813	32,872	39,134	41,811	44,536

Sedangkan hasil klusterisasi dengan menggunakan algoritma K-Means disajikan dalam tabel 5 berikut.

Tabel 5. Hasil Klusterisasi dengan Algoritma K-Means

Klaster	Anggota	Jumlah	Kategori
Klaster 1	DKI Jakarta dan DI Yogyakarta	2	Sangat Baik
Klaster 2	Jawa Barat, Jawa Tengah, Jawa Timur, Sumatera Barat, Kalimantan Selatan, Kalimantan Timur, Bali, Bangka Belitung, Banten, dan Kepulauan Riau	10	Baik
Klaster 3	Sumatera Utara, Riau, Jambi, Sumatera Selatan, Lampung, Kalimantan Barat, Kalimantan Tengah, Sulawesi Tenggara, Bengkulu, Papua Barat, dan Kalimantan Utara	11	Cukup
Klaster 4	Aceh, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Maluku, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua, Maluku Utara, Gorontalo, dan Sulawesi Barat	12	Kurang

Tahap selanjutnya dari proses *data mining* adalah evaluasi pola yang bertujuan mengidentifikasi pola relevan yang mampu mewakili pengetahuan berdasarkan pada beberapa tindakan yang tepat. Evaluasi dilakukan menggunakan *Partition Coefficient Index* (PCI) dan *Silhouette Coefficient* (SC). Bezdek (dalam Ramadhan dkk., 2019) mengusulkan validitas klusterisasi dengan mengevaluasi nilai keanggotaan data pada setiap klaster. Nilai *Partition Coefficient Index* (PCI) hanya mengevaluasi nilai derajat keanggotaan, tanpa memandang vektor (data) yang biasanya mengandung informasi geometric. Nilai dalam rentangan [0,1]. Pada pengujian PCI, klaster optimal didapat jika nilai yang diperoleh semakin besar (mendekati 1). Nilai PCI yang diperoleh adalah 0,81. Nilai ini mendekati 1, sehingga kualitas klaster dapat dikatakan baik. *Silhouette Coefficient* adalah sebuah teknik yang digunakan untuk mengukur seberapa baik letak objek dalam klaster dengan memberikan sebuah representasi grafis singkat (Gama dkk., 2014). Nilai *Silhouette Coefficient* (SC) yang diperoleh adalah 0,43 (*weak structure*). Nilai $SC > 0,25$ menunjukkan bahwa terdapat hubungan atau kemiripan antar data dalam suatu klaster. Kemiripan yang dimaksud dilihat berdasarkan jarak/kedekatan data.

Tahap terakhir dalam *data mining* adalah presentasi pengetahuan yaitu menyajikan atau memvisualisasikan pengetahuan atau informasi yang diperoleh. Presentasi pengetahuan yang dilakukan dalam penelitian ini adalah dengan memberikan kontribusi ilmiah berupa karya

ilmiah yang berjudul “Penerapan Algoritma K-Means pada Hasil Ujian Nasional SMA/MA di Indonesia”. Melalui skripsi ini, diharapkan bisa memberikan kontribusi ilmiah dalam menganalisa kualitas pendidikan SMA/MA masing-masing provinsi di Indonesia berdasarkan hasil UN serta dapat membantu melakukan evaluasi pendidikan demi terciptanya pendidikan yang lebih maju di masa depan.

Hasil kluster yang telah diperoleh adalah pengetahuan/informasi yang bermanfaat bagi pengguna kebijakan dalam proses pengambilan keputusan. Dengan adanya kluster tersebut diharapkan memberikan gambaran secara umum pencapaian kompetensi lulusan masing-masing provinsi berdasarkan nilai ujian nasional. Dari sini bisa diketahui wilayah mana saja yang memerlukan pembinaan mutu pendidikan, baik dari sisi peningkatan sumber daya manusia (Guru) di daerah/sekolah tersebut, sarana prasarana sekolah, pembangunan di sekolah, fasilitas belajar dan lain-lain. Untuk masing-masing instansi diimbangi dengan melaksanakan evaluasi sekolah dan mencari solusi bagaimana agar kualitas sekolah dapat terus meningkat. Guru juga harus lebih mengembangkan model pembelajaran, media pembelajaran dan lainnya agar dapat meningkatkan kemampuan kognitif siswa. Pada akhirnya peningkatan mutu pendidikan akan lebih terencana dengan baik karena telah dilakukan pemetaan kualitas.

PENUTUP

Berdasarkan hasil penelitian dapat disimpulkan bahwa Algoritma K-Means merupakan algoritma yang sederhana dan cepat untuk menyelesaikan masalah kluster terutama untuk data numerik serta sangat fleksibel dan efisien untuk ukuran data yang cukup besar dan menyebar (bervariasi). Dengan algoritma K-Means bisa ditentukan provinsi mana yang termasuk ke dalam kluster dengan kategori pencapaian sangat baik, baik, cukup, maupun kurang. Hasil kluster yang diperoleh adalah pengetahuan/informasi yang bermanfaat bagi pengguna kebijakan dalam proses pengambilan keputusan.

Berdasarkan hasil penelitian yang telah diperoleh, adapun saran yang dapat peneliti sampaikan, yaitu: (1) Kepada pemerintah disarankan agar melakukan evaluasi kepada provinsi yang dikategorikan tingkat pencapaiannya rendah serta memberikan bantuan dan pembinaan agar kualitas pendidikan di daerah tersebut dapat meningkat; (2) Kepada daerah yang tingkat pencapaian hasil UN dikategorikan rendah agar melaksanakan evaluasi pada tingkat yang lebih kecil misalnya di tingkat kabupaten, kecamatan, desa, maupun sekolah; dan (3) Kepada peneliti yang tertarik dengan penerapan algoritma K-Means disarankan melakukan penelitian untuk data lain seperti data ujian nasional tingkat kabupaten, tingkat kecamatan dan lain sebagainya, bahkan untuk data di luar bidang pendidikan.

DAFTAR PUSTAKA

- Aditya,A., Jovian, I. , Sari BN. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *Jurnal Media Informatika Budidarma* 4 (1), 51-58, 2020.
- Hartanti, NT. (2018). Education data mining untuk menentukan kelompok belajar ujian nasional di SMK. *Jurnal Tekno Kompak* 12 (2), 39-44, 2018.

- Irwan, B. (2012). Data Clustering Menggunakan Metodologi CRISP-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tri Dharma.
- Khomarudin, A. N. (2016). Teknik Data Mining: Alogaritma K-Means Clustering. *IlmuKomputer.com* , 3.
- Mulyani, EDS., Agustin, YH., Surgawi, NM., Susanto, S (2018). Implementasi Algoritma K-Means dan Fp-Growth untuk Rekomendasi Bimbingan Belajar Berdasarkan Segmentasi Akademik Siswa. *IT (Informatic Technique) Journal* 6 (2), 160-173, 2018.
- Prasetyo, E. (2014). Data Mining : Mengolah Data Menjadi Informasi Menggunakan Matlab.
- R.Mythily, A.Banu, & Raghunathan, S. (2015). Clustering Models for Data Stream Mining.
- Ramadhan, A., Mustakim, & Handinata, R. (2019). Implementasi Algoritma Fuzzy C-Means untuk Pengelompokan Wilayah Bencana Banjir. *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 11 Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau* , 173.
- Sutriyani, TP., Siregar, AM., Kusumaningrum, DS. (2018). Implementasi Algoritma K-Means Terhadap Pengelompokan Nilai Ujian Nasional Tingkat SMP di Provinsi Jawa Barat. *Techno Xplore: Jurnal Ilmu Komputer dan Teknologi Informasi* 3 (1), 30-36, 2018.
- Tohir, M. (2019). Hasil PISA Indonesia Tahun 2018 Turun dibandingkan Tahun 2015. 1.
- Yadav, R., & Sharma, A. (2012). Advance Methods to Improve Performance of K-Means Algorithm. *Global Journal of Computer Science and Technology* , 3.