

Implementasi Algoritma K-Means Terhadap Pengelompokan Nilai Ujian Nasional Tingkat SMP di Provinsi Jawa Barat

¹Tria Pratiwi Sutriyani, ²Amril Mutoi Siregar, ³Dwi Sulistya Kusumaningrum

¹if15.triasutriyani@mhs.ubpkarawang.ac.id, ²amrilmutoi@ubpkarawang.ac.id,

³dwi.sulistya@ubpkarawang.ac.id

Universitas Buana Perjuangan Karawang
Karawang, Indonesia

ABSTRAK

Peraturan Pemerintah Republik Indonesia Nomor 13 Tahun 2015 Pasal 66 menjelaskan bahwa Ujian Nasional (UN) sangatlah penting karena menjadi suatu pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu pada Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP) dan Sekolah Menengah Atas (SMA). Berdasarkan data yang ditemukan pada *website* Kementerian Pendidikan dan Kebudayaan yaitu nilai rata-rata UN SMP, SMA dan SMK data yang disajikan tidak berdasarkan peringkat atau kategori tertentu sehingga informasi yang didapatkan menjadi kurang jelas. Dari permasalahan tersebut dilakukan sebuah penelitian untuk menambang data menjadi informasi, pada penelitian ini data yang digunakan adalah nilai rata-rata UN tingkat SMP di Provinsi Jawa Barat menggunakan teknik *data mining*. Penelitian dilakukan dengan cara perhitungan manual menggunakan Microsoft Excel kemudian dievaluasi menggunakan *tools* Rapidminer 9.2.0. Penelitian ini menghasilkan *Cluster* Baik 10 Kabupaten/Kota, *Cluster* Cukup Baik 15 Kabupaten/Kota dan *Cluster* Kurang Baik 2 Kabupaten/Kota. Mata Pelajaran Matematika dan IPA saat ini tidak menentukan kategori, karena nilai yang didapat rata-rata lebih kecil dari Bahasa Inggris dan Bahasa Indonesia sehingga bagi Pemerintah Dinas Pendidikan untuk perlu meningkatkan lagi dalam proses pembelajaran di Provinsi Jawa Barat.

Kata kunci — *Algoritma K-Means, Clustering, Data Mining, UN SMP.*

I. PENDAHULUAN

Ujian Nasional atau UN merupakan salah satu penunjang keputusan untuk menyatakan lulus atau tidaknya siswa menurut Peraturan Pemerintah Republik Indonesia Nomor 13 Tahun 2015 Pasal 66 menyebutkan bahwa Penilaian hasil belajar sebagaimana dimaksud dalam Pasal 63 ayat (1) huruf c bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu dan dilakukan dalam bentuk UN. Selain itu, UN juga berpengaruh untuk melanjutkan jenjang pendidikan berikutnya, serta dapat meningkatkan mutu pendidikan sekolah tersebut sebagaimana tercantum pada Pasal 68 [1]. Kemudian ditemukan data nilai rata-rata UN Sekolah Menengah Pertama (SMP), Sekolah Menengah Atas (SMA) dan Sekolah Menengah Kejuruan (SMK) Tahun 2017 pada website Data Indonesia dalam Satu Portal [2] dengan rujukan dari *website* resmi Kementerian Pendidikan dan Kebudayaan (KEMDIKBUD) [3]. Data di publikasikan dalam bentuk format file Microsoft Excel (.xls) akan tetapi data yang disajikan masih secara acak tidak berdasarkan kategori atau peringkat tertentu sehingga informasi menjadi kurang jelas. Berdasarkan data yang diperoleh pada website KEMDIKBUD tersebut dilakukan penelitian menggunakan teknik *data mining* untuk menambang data menjadi informasi.

Terdapat banyak metode yang digunakan dalam *data mining* salah satunya adalah *clustering* atau pengelompokan merupakan metode yang digunakan untuk data yang tidak berlabel (*unlabeled data*) [4]. Salah satu algoritma dari metode *clustering* yaitu algoritma *clustering K-Means*. Berdasarkan X. Wu *et al.* [5] algoritma *clustering K-Means* merupakan sepuluh algoritma terbaik pada *data mining* yang digunakan untuk mengelompokkan data-data ke dalam *Cluster* berdasarkan suatu kemiripan variabel atau atribut data. Penelitian algoritma *K-Means* telah dibuktikan berhasil untuk mengelompokkan data oleh Yusuf dan Eka [6] tentang Penerapan Algoritma *K-Means clustering* pada Aplikasi menentukan Berat Badan Ideal dengan cara membandingkan titik pusat masing-masing data dengan titik pusat *cluster* yang telah ditetapkan. Selanjutnya pada penelitian Fitra, Abdul dan Mukis [7] tentang Pengelompokan Provinsi di Indonesia berdasarkan Karakteristik Kesejahteraan Rakyat menggunakan Metode *K-Means Cluster* yang membuktikan bahwa metode *K-Means cluster* dapat digunakan untuk pengelompokan Provinsi di Indonesia dibuat ke dalam tiga cluster. Hasil pengelompokan menunjukkan bahwa terdapat 6 provinsi pada cluster satu, 8 provinsi pada cluster dua, dan 19 provinsi pada cluster tiga. Kemudian, penelitian tentang Pengelompokan Mahasiswa berdasarkan

nilai Ujian Nasional dan IPK menggunakan Metode *K-Means* dilakukan oleh Hartatik [8] membuktikan bahwa metode *K-Means* dapat digunakan untuk melakukan pengelompokan mahasiswa berdasarkan dua parameter yaitu IPK dan nilai UN mahasiswa tersebut ketika lulus SMU. Hasil *cluster* yang didapatkan adalah $\{0 ; 0\}$ untuk *cluster* yang pertama, $\{2,611 ; 6,38\}$ untuk *cluster* yang kedua dan $\{3,44 ; 7,095\}$ untuk *cluster* yang ketiga.

Pada penelitian ini data yang diteliti yaitu nilai rata-rata UN di Provinsi Jawa Barat pada Sekolah Menengah Pertama dengan menggunakan atribut Mata Pelajaran Bahasa Indonesia, Bahasa Inggris, Matematika dan Ilmu Pengetahuan Alam (IPA). Setelah dilakukan penelitian nilai rata-rata UN tingkat SMP di Provinsi Jawa Barat diharapkan mampu menambah dan memperjelas informasi dalam menunjang keputusan terutama pada Dinas Pendidikan Provinsi Jawa Barat berdasarkan nilai rata-rata UN.

Berdasarkan latar belakang masalah yang telah dipaparkan, maka penelitian ini menjabarkan permasalahan menjadi beberapa masalah yaitu bagaimana menerapkan Algoritma *Clustering K-Means* untuk mengelompokkan Kabupaten dan Kota di Provinsi Jawa Barat berdasarkan Nilai UN tingkat Sekolah Menengah Pertama dan bagaimana hasil pengelompokan Kabupaten dan Kota di Provinsi Jawa Barat berdasarkan Nilai UN tingkat Sekolah Menengah Pertama dengan Algoritma *Clustering K-Means*.

II. STUDI LITERATUR

A. Ujian Nasional

Menurut Peraturan Pemerintah Republik Indonesia Nomor 13 Tahun 2015 Pasal 66 menyebutkan bahwa Penilaian hasil belajar sebagaimana dimaksud dalam Pasal 63 ayat (1) huruf c bertujuan untuk menilai pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu dan dilakukan dalam bentuk UN. UN dilakukan secara obyektif, berkeadilan, dan akuntabel. UN diadakan paling sedikit 1 (satu) kali dalam 1 (satu) tahun pelajaran. Hasil UN digunakan sebagai salah satu dasar untuk pemetaan mutu program dan/atau satuan pendidikan, pertimbangan seleksi masuk jenjang pendidikan berikutnya pembinaan dan pemberian bantuan kepada satuan pendidikan dalam upayanya untuk meningkatkan mutu pendidikan Pasal 68 [1].

Peserta Didik pendidikan informal dapat memperoleh ijazah yang setara dengan ijazah dari pendidikan dasar dan menengah jalur formal atau nonformal setelah lulus uji kompetensi yang diselenggarakan oleh satuan pendidikan yang terakreditasi dan telah mengikuti UN bagi Peserta Didik sederajat SMP/MTs atau SMA/MA/SMK/MAK. Kriteria pencapaian Standar Kompetensi Lulusan, untuk jenjang SMP/SMPLB/MTs atau bentuk lain yang sederajat, SMA/SMALB/MA atau bentuk lain yang sederajat, dan SMK/MAK atau bentuk lain yang sederajat, Menurut Peraturan Pemerintah Republik Indonesia Nomor 13 Tahun 2015 [1].

B. Data Mining

Data mining muncul sekitar tahun 90-an juga sering dikenal sebagai *Knowledge Discovery in Database (KDD)* [9]. Kehadirannya didasarkan dengan masalah membeludaknya data yang akhir-akhir ini dimana banyak organisasi telah mengumpulkan data sekian tahun lamanya (data pembelian, data penjualan dan sebagainya). *Data mining* mengeksplorasi *database* untuk menemukan pola-pola yang tidak terlihat seperti mencari informasi untuk memprediksi penjualan yang mungkin saja terlupakan oleh pebisnis. *Data mining* adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di *database* yang besar [10]. Lebih lanjut, menurut *Gartner Group data mining* adalah suatu proses menemukan kolerasi, pola, dan tren baru yang bermakna dengan memilah-milah sejumlah besar data yang tersimpan dalam penyimpanan data, menggunakan teknologi pengenalan pola serta teknik statistik dan matematika.

C. Pengelompokan Data Mining

Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu [9]:

1. Deskripsi
Deskripsi memberikan penjelasan untuk suatu pola atau kecenderungan.
2. Estimasi
Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi.
3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.

4. **Klasifikasi**
Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.
5. **Pengelompokan (*clustering*)**
Pengelompokan merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan *record* yang memiliki kemiripan suatu dengan yang lainnya dan memiliki ketidakmiripan dengan *record* dalam kluster lain. Pengelompokan berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengelompokan. Pengelompokan tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target.
6. **Asosiasi**
Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

D. Algoritma Clustering K-Means

Fokus penelitian ini adalah Algoritma *K-Means* merupakan salah satu algoritma *Clustering*, tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan berupa data tanpa label kelas [4]. Algoritma standar *K-Means* adalah Algoritma *Hartigan-Wong* yang didefinisikan pada persamaan (1).

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad (1)$$

Dimana :

- x_i : sebuah data milik Cluster C_k
 μ_k : nilai rata-rata (mean) dari data-data pada Cluster C_k

Langkah-langkah Algoritma *Clustering K-Means* dapat dijelaskan sebagai berikut :

1. Tentukan k buah *Cluster*.
2. Pilih sejumlah k buah objek secara acak yang akan dijadikan sebagai titik *centroid Cluster*.
3. Tentukan k buah *centroid* (titik tengah)
4. Kelompokkan objek ke *centroid Cluster* terdekat berdasarkan jarak objek ke *centroid* (*Euclidean distance*)
5. Hitung kembali semua nilai titik *centroid*
6. Ulangi langkah 3,4 dan 5 sampai nilai titik *centroid* tidak lagi berubah.

Selain *Hartigan-Wong*, terdapat rumus lainnya yaitu salah satunya adalah *Euclidian Distance* digunakan untuk menghitung titik *centroid* yang dijelaskan pada persamaan (2).

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Dimana :

- x_i : sebuah data milik Cluster C_k
 y_i : Nilai objek

E. Karakteristik K-Means

Adapun Karakteristik Algoritma *K-Means* adalah sebagai berikut [4]:

1. **Clustering**
Clustering adalah proses pengelompokan objek-objek ke dalam beberapa kelompok yang berbeda, atau lebih tepatnya berpartisipasi *dataset* menjadi *subsets*.
2. **Hierarchical Clustering**
Merupakan jenis *Clustering* yang kompleks atau disebut juga *multilevel hierarchy*. Cluster yang lebih besar dikelompokkan lagi menjadi dua atau lebih *Cluster* yang lebih kecil dan juga membentuk *tree diagram* yang disebut *dendogram*.
3. **Density-based Clustering**
Merupakan jenis *Clustering* yang berhubungan dengan kerapatan objek (densitas), dimana sekelompok objek yang lebih padat dipisahkan oleh sekelompok objek yang lebih renggang.
4. **Model-based Clustering**

Merupakan jenis *Clustering* yang objeknya dibentuk melalui model, misal menggunakan model statistik standar.

5. Dataset untuk *K-Means*

Dataset merupakan sekumpulan data yang digunakan oleh Algoritma *K-Means*, dibagi menjadi dua jenis yaitu data berlabel (*labeled data*) dan data tanpa label (*unlabeled data*). Algoritma *supervised learning* memerlukan data berlabel. Sedangkan *Unsupervised learning* memerlukan data tanpa label. *K-Means* tergolong pada *Unsupervised learning* maka memerlukan data tanpa label. Selain itu bentuk data juga harus berupa bilangan atau numerik.

6. Nilai K

Algoritma *K-Means* memerlukan nilai k, yang menyatakan jumlah *Cluster*, nilai K ini harus diinputkan diawal. Pilihan K yang tepat akan sangat mempengaruhi performa dari *K-Means*. Nilai k dapat ditebak atau menentukan nilai K sesuai kebutuhan.

7. *Centroid*

Setiap *Cluster* akan memiliki titik tengah (pusat) yang disebut *centroid*. Pertama *centroid* ditentukan secara acak, selanjutnya nilai *centroid* dijelaskan pada persamaan (3).

$$c = \sum_{i=1}^n \frac{x_i}{n} \quad (3)$$

Dimana :

C : *Centroid* pada *cluster*

Xi : Titik/objek ke – n

F. *Cross-Industry Standard Process for Data Mining (CRISP-DM)*

Cross-Industry Standard Process for Data Mining (CRISP-DM) yang dikembangkan tahun 1996 oleh analis dari beberapa *industry* seperti *Daimler Chrysler*, *SPSS*, dan *NRC* [11]. *CRISP-DM* menyediakan standar proses data mining sebagai strategi pemecah masalah secara umum dari bisnis dan unit penelitian. Dalam *CRISP-DM* sebuah proyek *data mining* memiliki siklus hidup yang terbagi dalam enam fase. Keseluruhan fase berurutan yang ada tersebut bersifat adaptif, fase berikutnya dalam urutan bergantung pada keluaran dari fase sebelumnya. Hubungan penting antarfase dihubungkan dengan tanda panah. Sebagai contoh jika fase berada dalam fase *modelling*. Berdasarkan pada perilaku dan karakteristik model, proses mungkin harus kembali kepada fase *data preparation* untuk perbaikan lebih lanjut terhadap data atau berpindah maju kepada fase *evaluation*.

G. *Rapidminer*

Rapidminer adalah perangkat lunak yang dapat diakses secara bebas (*open source*) untuk melakukan analisis *data mining*, *text mining*, dan analisis prediksi [10]. Perangkat lunak ini memiliki unggulan yakni berbasis *open source* di seluruh dunia yang termuka dan ternama serta lebih handal dalam menyediakan pembelajaran skema, model dan algoritma dari *Weka* dan *R Script*, selain *data mining* *Rapidminer* juga menyediakan fungsi seperti precessing data dan visualisasi, analisis prediktif dan pemodelan statistik, evaluasi dan penyebaran

III. IMPLEMENTASI ALGORITMA K-MEANS TERHADAP PENGELOMPOKAN NILAI UJIAN NASIONAL TINGKAT SMP DI PROVINSI JAWA BARAT

A. *Tahap Pencarian Dataset*

Tahap pencarian dataset dilakukan dengan mencari data website di Data Indonesia dalam Satu Portal bersumber dari Kementerian Pendidikan dan Kebudayaan (KEMDIKBUD) [2].Data yang didapatkan berjumlah 2,196 data 14 atribut sehingga dilakukan penyeleksian data.

B. *Tahap Seleksi Data*

Dari data yang berjumlah 2,196 yang berisi 35 Provinsi dan 514 Kabupaten atau Kota lalu dari masing-masing data tersebut terdapat Sekolah Menengah Pertama, Sekolah menengah Atas Jurusan IPA dan IPS serta Sekolah Menengah Kejuruan dengan total keseluruhan yakni 2,196 data. Selain jumlah data tersebut terdapat juga 13 atribut yakni nilai rata-rata UN Mata Pelajaran Bahasa Indonesia, Bahasa Inggris, Matematika, IPA, Fisika, Kimia, Biologi, Ekonomi, Sosiologi, Kejuruan, Jumlah Peserta, *Latitude*, *Longitude*. Berdasarkan data tersebut dilakukan penyeleksian data untuk memfokuskan penelitian dengan mengambil Data Sekolah Menengah Pertama di

Provinsi Jawa Barat yang berjumlah 27 Kabupaten atau Kota dan Empat atribut yaitu data nilai rata-rata UN. Data yang telah diseleksi terdapat pada Tabel 1.

Tabel 1. Data yang akan diproses untuk pengelompokan

No	Kode	Kabupaten/Kota	Bahasa Indonesia	Bahasa Inggris	MTK	IPA
1	1.0	Kab. Bandung	66.85	55.34	53.31	55.93
2	2.0	Kab. Bandung Barat	66.45	54.61	53.71	55.90
3	3.0	Kab. Bekasi	65.74	53.69	51.08	54.38
4	4.0	Kab. Bogor	67.11	54.80	52.65	55.41
5	5.0	Kab. Ciamis	70.54	56.72	55.89	57.59
6	6.0	Kab. Cianjur	68.11	62.81	61.48	62.91
7	7.0	Kab. Cirebon	66.21	51.96	50.67	53.70
8	8.0	Kab. Garut	65.11	52.88	52.03	54.66
9	9.0	Kab. Indramayu	64.02	51.29	50.09	53.46
10	10.0	Kab. Karawang	65.26	56.84	56.21	58.23
11	11.0	Kab. Kuningan	69.84	53.28	51.07	54.36
12	12.0	Kab. Majalengka	68.28	53.83	52.00	54.34
13	13.0	Kab. Pangandaran	66.41	51.65	50.17	53.10
14	14.0	Kab. Purwakarta	65.67	51.52	49.97	53.22
15	15.0	Kab. Subang	65.20	46.14	41.10	45.54
16	16.0	Kab. Sukabumi	65.53	54.07	52.96	55.20
17	17.0	Kab. Sumedang	67.82	48.90	42.22	46.53
18	18.0	Kab. Tasikmalaya	67.59	54.48	53.04	55.68
19	19.0	Kota Bandung	75.04	62.86	51.73	57.44
20	20.0	Kota Banjar	69.60	52.70	50.04	53.68
21	21.0	Kota Bekasi	73.02	61.23	55.01	57.82
22	22.0	Kota Bogor	73.72	63.65	59.12	61.79
23	23.0	Kota Cimahi	74.55	60.48	55.40	58.23
24	24.0	Kota Cirebon	73.04	61.04	57.05	59.16
25	25.0	Kota Depok	72.33	59.04	53.43	56.89
26	26.0	Kota Sukabumi	72.09	59.06	54.64	57.63
27	27.0	Kota Tasikmalaya	72.44	63.08	60.04	63.67

C. Tahap Implementasi Algoritma Clustering K-Means

Setelah proses penyeleksian data lalu tahap selanjutnya yaitu pengolahan *dataset* menggunakan Algoritma *Clustering K-Means* dengan menentukan *Cluster*. Saat ini belum ada indikator resmi untuk menentukan kategori sehingga pada penelitian ini digunakan tiga *Cluster* yaitu Baik, Cukup Baik dan Kurang Baik.

D. Hasil Pengelompokan

Hasil pengelompokan ini nantinya menghasilkan Kabupaten atau Kota mana saja yang merupakan *Cluster* Baik, Cukup Baik dan Kurang Baik berdasarkan nilai rata-rata UN

E. Tahap Evaluasi

Setelah mengetahui hasil pengelompokan nilai rata-rata UN tingkat SMP di Provinsi Jawa Barat secara perhitungan manual menggunakan Microsoft Office dengan Algoritma *K-Means*, selanjutnya dilakukan evaluasi menggunakan *tools* Rapidminer untuk membandingkan hasil keduanya.

IV. HASIL DAN PEMBAHASAN

Hasil Perhitungan manual menggunakan Microsoft Excel 2013 dilakukan sebanyak 8 iterasi sampai titik *centroid* tidak lagi berubah dengan menghitung jarak menggunakan persamaan (2). Setelah menghitung jarak kemudian dilakukan perhitungan titik *centroid* baru dengan menggunakan persamaan (3). Sehingga hasil iterasi terakhir didapat pada Tabel 2 dimana angka yang bercetak tebal (*bold*) berada pada *cluster* C1, C2 atau C3.

Tabel 2. Hasil Pengelompokan secara Perhitungan Manual

Kabupaten/Kota	Bahasa Indonesia	Bahasa Inggris	MTK	IPA	C1	C2	C3
Kab. Bandung	66.85	55.34	53.31	55.93	17.17270	9.20066	2.57066
Kab. Bandung Barat	66.45	54.61	53.71	55.90	17.11118	9.79915	2.38428
Kab. Bekasi	65.74	53.69	51.08	54.38	14.03699	12.30210	1.30741
Kab. Bogor	67.11	54.80	52.65	55.41	16.18731	9.82164	1.66744
Kab. Ciamis	70.54	56.72	55.89	57.59	20.90196	5.02944	7.00534

Kabupaten/Kota	Bahasa Indonesia	Bahasa Inggris	MTK	IPA	C1	C2	C3
Kab. Cianjur	68.11	62.81	61.48	62.91	30.23148	7.83661	15.68446
Kab. Cirebon	66.21	51.96	50.67	53.70	12.63867	13.60805	2.31306
Kab. Garut	65.11	52.88	52.03	54.66	14.58140	12.68404	1.67161
Kab. Indramayu	64.02	51.29	50.09	53.46	12.10820	15.48360	4.11892
Kab. Karawang	65.26	56.84	56.21	58.23	21.18597	8.40988	6.57861
Kab. Kuningan	69.84	53.28	51.07	54.36	14.21690	10.92162	3.34212
Kab. Majalengka	68.28	53.83	52.00	54.34	14.79316	10.62918	1.71384
Kab. Pangandaran	66.41	51.65	50.17	53.10	11.80683	14.19371	3.06901
Kab. Purwakarta	65.67	51.52	49.97	53.22	11.72117	14.64577	3.34410
Kab. Subang	65.20	46.14	41.10	45.54	2.04429	26.40054	16.08958
Kab. Sukabumi	65.53	54.07	52.96	55.20	15.98594	11.18049	1.67073
Kab. Sumedang	67.82	48.90	42.22	46.53	2.04429	23.06534	13.59152
Kab. Tasikmalaya	67.59	54.48	53.04	55.68	16.49662	9.52939	1.97156
Kota Bandung	75.04	62.86	51.73	57.44	23.22833	5.92493	12.83824
Kota Banjar	69.60	52.70	50.04	53.68	12.84721	12.21042	3.76183
Kota Bekasi	73.02	61.23	55.01	57.82	23.39771	2.10725	10.90119
Kota Bogor	73.72	63.65	59.12	61.79	29.41484	4.71220	15.92790
Kota Cimahi	74.55	60.48	55.40	58.23	23.87695	2.57885	11.61612
Kota Cirebon	73.04	61.04	57.05	59.16	25.19026	0.88329	11.95570
Kota Depok	72.33	59.04	53.43	56.89	20.56567	4.29213	8.33727
Kota Sukabumi	72.09	59.06	54.64	57.63	21.61556	3.12522	8.70979
Kota Tasikmalaya	72.44	63.08	60.04	63.67	30.43183	6.06033	16.43428

Setelah mengetahui hasil pengelompokan pada Tabel 2, selanjutnya menentukan kategori Baik, Cukup Baik dan Kurang Baik dengan mengurutkan jumlah nilai dari terkecil, sedang dan tertinggi. Sehingga didapatkan hasil akhir pengelompokan secara manual yang terdapat pada Tabel 3.

Tabel 3. Hasil Pengelompokan Manual

Cluster	Kabupaten/Kota	Jumlah
Baik	Kabupaten Ciamis, Kabupaten Cianjur, Kota Bandung, Kota Bekasi, Kota Bogor, Kota Cimahi, Kota Cirebon, Kota Depok, Kota Sukabumi, Kota Tasikmalaya	10
Cukup Baik	Kabupaten Bandung, Kabupaten Bandung Barat, Kabupaten Bekasi, Kabupaten Bogor, Kabupaten Cirebon, Kabupaten Garut, Kabupaten Indramayu, Kabupaten Karawang, Kabupaten Kuningan, Kabupaten Majalengka, Kabupaten Pangandaran, Kabupaten Purwakarta, Kabupaten Sukabumi, Kabupaten Tasikmalaya, Kota Banjar	15
Kurang Baik	Kabupaten Subang, Kabupaten Sumedang	2

Berdasarkan hasil perhitungan secara manual menggunakan Microsoft Excel pada Tabel 2 dan otomatis menggunakan *tools* Rapidminer pada Tabel 4. Kelompok yang dihasilkan sama, yaitu Cluster Baik 10 Kabupaten/Kota, Cluster Cukup Baik 15 Kabupaten/Kota dan Kurang Baik 2 Kabupaten/Kota.

Tabel 4. Hasil Pengelompokan *tools* Rapidminer

Cluster	Kabupaten/Kota	Jumlah
Baik	5.0, 6.0, 19.0, 21.0, 22.0, 23.0, 24.0, 25.0, 26.0, 27.0	10
Cukup Baik	1.0, 2.0, 3.0, 4.0, 7.0, 8.0, 9.0, 10.0, 11.0, 12.0, 13.0, 14.0, 16.0, 18.0, 20.0	15
Kurang Baik	15.0, 17.0	2

Hasil pengelompokan Kabupaten dan Kota di Provinsi Jawa Barat berdasarkan nilai UN tingkat SMP dibagi menjadi tiga kelompok yaitu Baik sebanyak 15 Kabupaten/Kota, Cukup Baik sebanyak 10 Kabupaten/Kota dan Kurang Baik sebanyak 2 Kabupaten. Hasil tersebut didapat dari perhitungan manual menggunakan Microsoft Excel sebanyak 8 iterasi sampai titik *centroid* yang didapat tidak lagi berubah. Hasil yang didapat pada iterasi ke-8 titik *centroid* yang didapat sama maka proses iterasi dihentikan sampai dengan iterasi 8 walaupun ada beberapa jumlah anggota pada *cluster* yang sama di iterasi tiga dan lima, empat dan enam tidak berpengaruh karena titik *centroid* berbeda. Kabupaten atau Kota yang termasuk dalam kategori Baik memiliki jumlah nilai UN 233.45 – 263.84, kategori Cukup Baik 218.38 – 241.12 dan Kurang Baik 197.98 – 205.47. Penentuan kategori ditentukan oleh mata pelajaran Bahasa Inggris dan bahasa Indonesia karena dilihat dari rata-rata nilai UN mata pelajaran Bahasa Indonesia dan Bahasa Inggris cenderung lebih besar dari Matematika dan IPA sehingga Bahasa Indonesia dan bahasa Inggris sebagai penentu

dalam pengelompokan Kabupaten/Kota di Provinsi Jawa Barat berdasarkan nilai Ujian Nasional tingkat SMP ini.

V. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, maka kesimpulan dari penelitian ini adalah :

1. Algoritma *clustering K-Means* dapat digunakan untuk pengelompokan nilai UN tingkat SMP di Provinsi Jawa Barat dengan cara melakukan proses awal menyeleksi data, selanjutnya menerapkan algoritma *K-Means* yang diawali dengan menentukan nilai k atau jumlah *cluster* lalu menghitung jarak objek ke titik *centroid* kemudian hasil perhitungan jarak tersebut dibandingkan untuk melihat jarak terdekat setelah itu menghitung kembali titik *centroid* baru dan dilakukan secara berulang hingga titik *centroid* lama dan titik *centroid* baru tidak lagi berubah.
2. Hasil Pengelompokan menggunakan Algoritma *clustering K-Means* secara manual dan menggunakan *tools* menghasilkan kategori baik terdiri dari Kabupaten Ciamis, Kabupaten Cianjur, Kota Bandung, Kota Bekasi, Kota Bogor, Kota Cimahi, Kota Cirebon, Kota Depok, Kota Sukabumi, Kota Tasikmalaya. Kategori cukup baik terdiri dari Kabupaten Bandung, Kabupaten Bandung Barat, Kabupaten Bekasi, Kabupaten Bogor, Kabupaten Cirebon, Kabupaten Garut, Kabupaten Indramayu, Kabupaten Karawang, Kabupaten Kuningan, Kabupaten Majalengka, Kabupaten Pangandaran, Kabupaten Purwakarta, Kabupaten Sukabumi, Kabupaten Tasikmalaya, Kota Banjar dan kategori kurang baik terdiri dari Kabupaten Subang, Kabupaten Sumedang. Sehingga jika dikelompokkan menurut jumlah nilai UN jika Jumlah nilai UN 233.45 – 263.84 termasuk kategori baik, 218.38 – 214.12 termasuk kategori cukup baik dan 197.98 – 205.47 termasuk kategori kurang baik.

Hasil penelitian untuk meningkatkan Kategori Kabupaten/Kota di Provinsi Jawa Barat, nilai yang harus ditingkatkan adalah Bahasa Inggris dan Bahasa Indonesia, karena kedua mata pelajaran tersebut merupakan penentu untuk mendapatkan kategori Baik, Cukup Baik atau Kurang Baik sedangkan Matematika dan IPA tidak menjadi penentu karena kedua mata pelajaran tersebut hampir sama rata. Kemudian untuk penelitian selanjutnya diharapkan dapat menggunakan lebih dari tiga *cluster* dan dalam pemilihan titik *centroid* yang berbeda serta menggunakan algoritma *clustering* lainnya.

DAFTAR PUSTAKA

- [1] “Peraturan Pemerintah Republik Indonesia Nomor 13 Tahun 2015,” no. 2, 2015.
- [2] “Data Indonesia dalam Satu Portal,” 2017. .
- [3] “Jendela Pendidikan dan Kebudayaan.” .
- [4] R. Primartha, *Belajar Mechine Learning*. Bandung: Informatika Bandung, 2018.
- [5] X. Wu *et al.*, *Top 10 algorithms in data mining*. Verlag London Limiter, 2008.
- [6] Y. R. Nasution *et al.*, “Penerapan Algoritma K-Means Clustering pada Aplikasi menentukan Berat Badan Ideal,” vol. 6341, no. April, pp. 77–81, 2018.
- [7] J. Gaussian, “Pengelompokan Provinsi di Indonesia berdasarkan Karakteristik Kesejahteraan Rakyat menggunakan Metode K-Means Cluster,” vol. 4, no. 2006, pp. 875–884, 2015.
- [8] Hartatik, “Pengelompokan Mahasiswa Berdasarkan Nilai Ujian Nasional dan IPK menggunakan Metode K-Means,” pp. 35–40, 2014.
- [9] D. Larose, *Discovering Knowledge in Data*. Canada: A John Wiley & Sons, Inc, 2005.
- [10] A. Siregar and A. Paspabhuana, *Pengolahan Data menjadi Informasi dengan Rapidminer*. Surakarta, 2017.
- [11] Kusrini and E. Luthfi, *Algoritma data Mining*. Yogyakarta: ANDI, 2009.