

## *K-Means Clustering* untuk Mengelompokan Tingkat Putus Sekolah Jenjang SMP di Indonesia

Sekar Wuni  
Universitas Buana Perjuangan  
Karawang  
Karawang, Indonesia  
if15.sekarwuni@mhs.ubpkarawang.ac.id

Amril Mutoi Siregar  
Universitas Buana Perjuangan  
Karawang  
Karawang, Indonesia  
amril.mutoi99@gmail.com

Dwi Sulistya Kusumaningrum  
Universitas Buana Perjuangan  
Karawang  
Karawang, Indonesia  
dwi.sulistya@ubpkarawang.ac.id

### Abstrak—

Mendapatkan pendidikan adalah hak setiap warga Negara Indonesia, dari tingkat Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), hingga Sekolah Menengah Atas (SMA) maupun yang sederajat. Namun, ternyata masih ada siswa yang tidak dapat menyelesaikan pendidikannya pada tingkat tertentu maupun melanjutkan ke jenjang selanjutnya. Siswa yang tidak dapat menyelesaikan pendidikannya diberikan sebutan putus sekolah. Pada situs Kementerian Pendidikan dan Kebudayaan (Kemendikbud) terdapat banyak data yang disajikan, salah satunya adalah data angka putus sekolah dari berbagai jenjang pendidikan dan tahun ajar. Data-data tersebut belum jelas dikarenakan belum dapat dikategorikan maupun dikelompokkan berdasarkan besar kecilnya angka putus sekolah. Berdasarkan permasalahan tersebut, maka penelitian ini mengolah data angka putus sekolah jenjang SMP di Indonesia dengan menggunakan teknik pada *data mining* yaitu *clustering* (pengelompokan) dengan menggunakan algoritma *K-Means*. Pengelompokan pada data tersebut dibagi menjadi tiga kelompok yaitu kelompok tingkat putus sekolah jenjang SMP rendah, sedang, dan tinggi. Penelitian ini dilakukan menggunakan dua cara yaitu: perhitungan *manual* dengan menggunakan *Microsoft Excel 2013* dan pengujian menggunakan *tools Rapid Miner* versi 5.3.000. Hasil dari penelitian ini yaitu provinsi yang masuk pada kelompok tingkat putus sekolah jenjang SMP rendah sebanyak 21 provinsi, kelompok sedang sebanyak 12 provinsi dan yang masuk pada kelompok tinggi sebanyak 1 provinsi.

**Kata kunci** — *Clustering, Data mining, K-Means, Putus Sekolah, Sekolah Menengah Pertama (SMP).*

### I. PENDAHULUAN

Negara Indonesia telah mengatur perihal pendidikan dalam UUD 1945 pasal 31 ayat (1) yang berbunyi “Setiap warga negara berhak mendapatkan pendidikan” [1]. Mulai dari tingkat Sekolah Dasar (SD) maupun yang sederajat, Sekolah Menengah Pertama (SMP) maupun yang sederajat, Sekolah Menengah Atas (SMA) maupun yang sederajat, hingga jenjang Perguruan Tinggi (PT). Namun tidak setiap anak dapat menikmati pendidikan atau dapat menyelesaikan pendidikannya ke jenjang selanjutnya. Bagi peserta didik yang tidak mampu menyelesaikan suatu jenjang pendidikan formal, sehingga tidak dapat melanjutkan studinya ke jenjang pendidikan selanjutnya diberi predikat putus sekolah [2]. Lebih lanjut definisi dari putus sekolah adalah meninggalkan sekolah sebelum dapat menyelesaikan keseluruhan masa belajar yang telah ditetapkan oleh sekolah yang bersangkutan [2]. Hal ini dapat dilihat pada grafik perkembangan jumlah siswa putus sekolah menurut jenjang pendidikan dari tahun ajar 2015/2016 sampai dengan tahun ajar 2017/2018 yang diperoleh dari situs Kemendikbud [3].

Pada situs Kemendikbud [3] data angka putus sekolah untuk jenjang SD dan SMA cenderung mengalami penurunan dari tahun ajar 2015/2016 sampai dengan tahun ajar 2017/2018. Sedangkan, data angka putus sekolah jenjang SMP dan SMK yang mengalami penurunan pada tahun ajar 2016/2017, namun mengalami kenaikan pada tahun ajar 2017/2018. Kenaikan angka putus sekolah tahun ajar 2017/2018 pada jenjang SMP adalah sebanyak 12,488 siswa dan untuk jenjang SMK mengalami kenaikan sebanyak 644 siswa. Lebih lanjut dalam situs Kemendikbud [4] diperoleh data yang lebih rinci mengenai jumlah siswa SMP yang putus sekolah pada tingkat VII, VIII, dan IX di 34 provinsi di Indonesia. Selain itu pada situs yang tersebut, diperoleh data banyaknya siswa yang belajar pada jenjang SMP pada tahun ajar yang sama pada tingkat/kelas VII, VIII, dan IX di 34 provinsi di Indonesia. Data yang digunakan dalam penelitian ini adalah jumlah putus sekolah menurut tingkat tahun tertentu dibagi dengan jumlah siswa menurut tingkat tahun sebelumnya dikalikan dengan 100.

Data persentase angka putus sekolah jenjang SMP tersebut kemudian akan diolah dengan menggunakan teknik yang ada pada *data mining* yaitu *clustering* dengan menggunakan algoritma *K-Means*. *Data mining* merupakan salah satu cabang ilmu yang digunakan untuk menambang atau mengekstrak pengetahuan dari data yang jumlahnya banyak agar menjadi suatu informasi yang berguna dan dapat digunakan sebagai pendukung pengambilan keputusan [5] [6]. Penelitian sebelumnya dengan menggunakan algoritma *K-Means Clustering* dan telah terbukti berhasil, digunakan sebagai acuan dalam penyusunan penelitian. Adapun penelitian tersebut, diantaranya dilakukan oleh Kevin Bima Aditya *et al* [7], Agus Perdana Windarto [8], dan Joko Suryanto [9]. Kevin Bima Aditya *et al* [7] telah melakukan

penelitian tentang Sistem Informasi Geografis Pemetaan Faktor-Faktor yang Mempengaruhi Angka Kematian Ibu (AKI) dan Angka Kematian Bayi (AKB) dengan Metode *K-Means Clustering*. Hasil dari penelitian tersebut adalah 15% Kota/Kab. berada di tingkat rendah, 65% berada di tingkat sedang dan 20% berada di tingkat tinggi untuk persentase AKI-nya. Sedangkan persentase AKB-nya 32,5% Kota/Kab berada di tingkat rendah, 60% berada di tingkat sedang dan 7,5% berada di tingkat tinggi. Selanjutnya Agus Perdana Windarto [8] telah melakukan penelitian tentang *Implementation of Data mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method*. Penelitian tersebut menggunakan tiga kelompok untuk memetakan negara yang masuk dalam kategori negara dengan impor beras rendah, sedang, dan tinggi. Hasil penelitian tersebut adalah negara yang masuk pada kategori impor beras tinggi yaitu Vietnam dan Thailand. Kemudian Joko Suryanto [9] dengan penelitian tentang Analisa Perbandingan Pengelompokan Curah Hujan 15 Harian Provinsi DIY Menggunakan *Fuzzy Clustering* dan *K-Means Clustering*. Hasil dari penelitian tersebut menunjukkan bahwa algoritma *K-Means* lebih sesuai untuk pengelompokan curah hujan 15 harian Provinsi DIY dibandingkan metode *Fuzzy C-Means*.

Berdasarkan penelitian-penelitian yang telah dipaparkan, penelitian ini dimaksudkan untuk mengelompokkan tingkat putus sekolah jenjang SMP di Indonesia menjadi tiga kelompok yaitu kelompok tingkat putus sekolah jenjang SMP rendah, sedang dan tinggi. Metode yang digunakan adalah *clustering* dengan algoritma *K-Means*. Tujuan dari penelitian ini adalah untuk mengetahui provinsi di Indonesia yang masuk pada kategori kelompok putus sekolah jenjang SMP rendah, sedang, dan tinggi.

## II. STUDI LITERATUR

### A. Data mining

*Data mining* adalah suatu proses penggalian data yang belum pernah diketahui sebelumnya, namun dapat dipahami dan berguna dari basis data yang begitu besar serta digunakan untuk membuat keputusan bisnis yang penting [10]. Secara sederhana *data mining* disebut juga dengan *Knowledge Discovery From Data* (KDD) [11]. Dalam *data mining* ada beberapa jenis pengelompokan diantaranya berdasarkan pekerjaannya, berdasarkan fungsinya, dan berdasarkan tekniknya. Berdasarkan Pekerjaannya *data mining* dibagi menjadi empat kelompok yaitu: Model Prediksi, Analisis Kelompok, Analisis Asosiasi, dan Deteksi Anomali [12]. Berdasarkan Fungsinya *data mining* dibagi menjadi empat yaitu: Fungsi Prediksi (*prediction*), Fungsi Deskripsi (*description*), Fungsi Klasifikasi (*classification*), dan Fungsi Asosiasi (*asociation*) [13]. Berdasarkan Teknologinya ada tiga jenis teknik *data mining* yang populer yaitu: *Association Rule Mining*, *Clasification*, dan *Clustering* [10]. *Clustering* adalah suatu proses pengelompokan data tanpa harus menentukan kelas data tertentu terlebih dahulu. Prinsip dari *clustering* adalah memaksimalkan kesamaan data pada satu kelompok dan meminimalkan persamaan data antar kelompok. Ada tiga jenis *clustering* yaitu *Hierarki*, *Non-hierarki*, dan Gabungan dari *hierarki* dan *non-hierarki* (*hybrid*) [14]. Metode *clustering non-hierarki* adalah metode yang digunakan untuk mengelompokkan objek-objek pengamatan menjadi dua atau lebih kelompok [15].

### B. Algoritma *K-Means Clustering*

*K-Means* merupakan suatu metode pengelompokan berbasis jarak yang mempartisi data ke dalam dua atau lebih kelompok berdasarkan kesamaan karakteristiknya dan algoritma ini hanya bekerja untuk atribut angka [16] [7]. Setiap data yang ada harus masuk dalam kelompok tertentu pada suatu tahapan proses, namun pada tahapan proses berikutnya dapat berpindah pada kelompok yang lain [17]. Algoritma *K-Means* adalah salah satu algoritma yang paling populer dan yang paling cepat dalam teknik pengelompokan data serta bersifat *unsupervised*, dan mempunyai ketelitian yang cukup tinggi terhadap ukuran objek, sehingga relatif lebih terukur dan efisien untuk pengolahan data dengan jumlah yang relatif banyak serta tidak terpengaruh terhadap urutan objek [18] [19] [20]. *K-Means* adalah salah satu algoritma *clustering* yang masuk dalam metode *non-hierarki* atau *partial*, algoritma lain yang termasuk dalam metode *non-hierarki* diantaranya *Bisecting K-Mean*, *K-Mode*, *Partitioning Around Medoids* (PAM), *Clustering Large Applications* (CLARA), *Clustering Large Applications based on Range Search* (CLARANS), *Fuzzy C-Means* [21]. Pada dasarnya algoritma *K-Means* melakukan dua proses yaitu proses pendeteksian lokasi pusat kelompok (*centroid*) dan proses pencarian anggota dari masing-masing kelompok [22]. Adapun proses dasar algoritma *K-Means* seperti di bawah ini:

1. Tentukan  $k$  sebagai banyaknya *cluster* yang akan dibentuk dan tentukan pusat *cluster*.

- Menghitung jarak setiap data ke pusat *cluster* dengan menggunakan persamaan *Euclidean*, yang ditunjukkan pada persamaan (1).

$$d(i, k) = \sqrt{\sum_i^m (C_{ij} - C_{kj})^2} \tag{1}$$

- Mengelompokan data ke dalam *cluster* dengan jarak yang paling pendek dengan persamaan (2).

$$\min \sum_k^i -a_{ik} = \sqrt{\sum_i^m (C_{ij} - C_{kj})^2} \tag{2}$$

- Menghitung pusat *cluster* yang baru menggunakan persamaan (3).

$$C_{kj} = \frac{\sum_k^i x_{ij}}{p} \tag{3}$$

dengan  $X_{ij} \in cluster$  ke k p = banyaknya anggota *cluster* ke k

- Ulangi langkah ke-2 sampai dengan ke-4 sampai tidak ada data yang pindah ke *cluster* yang lain [23].

C. *Rapid Miner*

Siregar dan Puspabhuana [10] menjelaskan bahwa *Rapid Miner* merupakan sebuah perangkat lunak yang bersifat *open source* (dapat diunduh secara gratis). *Rapid Miner* dapat digunakan untuk melakukan analisis terhadap *data mining*, *text mining* dan dapat juga melakukan analisis prediksi, aplikasi ini memiliki kurang lebih 500 operator termasuk di dalamnya operator *input*, *output*, *data processing*, dan visualisasi. Teknik yang digunakan adalah berbagai macam teknik deskriptif dan prediksi. *Rapid Miner* ditulis dengan menggunakan bahasa pemrograman *java*, sehingga dapat bekerja di semua sistem operasi.

III. METODE PENELITIAN

Dalam penelitian ini metode penelitian yang digunakan yaitu: (a) Tahap pengumpulan data, (b) Tahap seleksi dan inisial data, (c) Tahap *K-Means Clustering*, (d) Tahap *performance* dan (e) Tahap analisis.

A. Tahap Pengumpulan Data

Adapun *dataset* yang akan digunakan adalah data angka putus sekolah jenjang SMP di Indonesia tahun ajar 2017/2018 yang diperoleh dari situs resmi Kemendikbud [4]. *Dataset* terdiri atas No. yang terdiri dari angka 1 sampai dengan 34, Provinsi yang terdiri dari nama 34 provinsi di Indonesia, tingkat yang terdiri dari VII, VIII, dan IX, dan terakhir adalah rata-rata seperti pada Tabel 1.

Tabel 1 Persentase Angka Putus Sekolah Menurut Tingkat Tiap Provinsi

No.	Provinsi <i>Province</i>	Tingkat/Grade			Rata-rata <i>Average</i>
		VII	VIII	IX	
1	DKI Jakarta	0.29	0.33	0.81	0.48
2	Jawa Barat	0.27	0.32	0.99	0.52
3	Banten	0.37	0.37	1.16	0.63
.					
.					
.					
32	Nusa Tenggara Timur	0.32	0.49	1.45	0.73
33	Papua	0.35	0.35	3.10	1.27
34	Papua Barat	0.41	0.41	1.51	0.84

B. Tahap Seleksi dan Inisial Data

Selanjutnya, data atau atribut yang tidak digunakan kemudian dibuang atau dihapus agar data dapat lebih mudah diolah. Data nama provinsi pada Tabel 1 yang berupa karakter kemudian diinisialkan menjadi angka seperti pada Tabel 2, hal ini dilakukan agar mempermudah pada saat proses perhitungan data secara *manual* menggunakan *Microsoft Excel 2013* maupun menggunakan *tools Rapid Miner* versi 5.3.000.

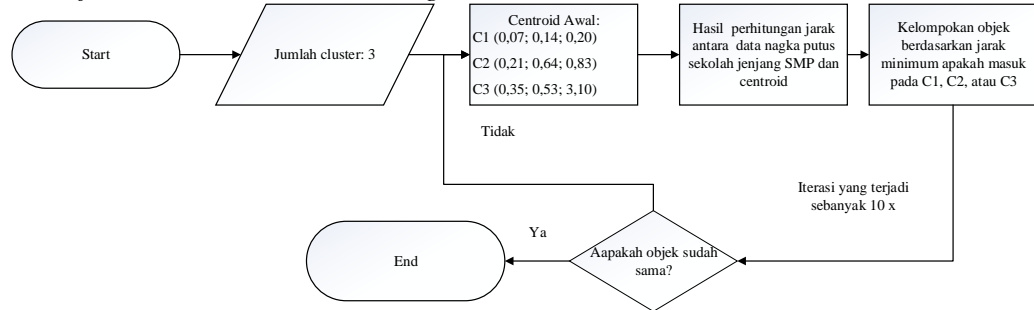
Tabel 2 *Dataset* setelah seleksi dan inisial

Provinsi	Tingkat		
	VII	VIII	IX
1	0.29	0.33	0.81
2	0.27	0.32	0.99
3	0.37	0.37	1.16

Provinsi	Tingkat		
	VII	VIII	IX
.			
.			
32	0.32	0.49	1.45
33	0.35	0.53	3.10
34	0.41	0.66	1.51

C. Tahap *K-Means Clustering*

Berikut adalah *flowchart K-Means Clustering*:



Gambar 1 *Flowchart K-Means Clustering*

Proses ini diawali dengan menentukan jumlah *cluster* yang akan dibentuk, pada penelitian ini *cluster* yang akan dibentuk ada tiga *cluster* yaitu C1, C2, dan C3. Setelah itu menentukan titik pusat *cluster* (*centroid*) untuk masing-masing *cluster*. Dalam penelitian ini penentuan *centroid* awal seperti pada Tabel 3.

Tabel 3 *Centroid awal*

Pusat Cluster	Tingkat		
	VII	VIII	IX
C1	0.07	0.14	0.20
C2	0.21	0.64	0.83
C3	0.35	0.53	3.10

Selanjutnya adalah menghitung jarak setiap data ke pusat *cluster* (*centroid*) dengan menggunakan rumus pada persamaan (1) dan persamaan (2). Setelah melakukan perhitungan, langkah selanjutnya adalah mengelompokan anggota *cluster* berdasarkan jarak terpendek. Selanjutnya adalah menghitung pusat *cluster* baru dengan cara jumlah data yang ada pada masing-masing *cluster* dibagi dengan banyaknya anggota *cluster*, sehingga diperoleh pusat *cluster* yang baru. Apabila pada iterasi selanjutnya anggota *cluster* tidak berpindah *cluster* maka proses dihentikan, apabila masih berpindah *cluster* maka proses dilanjutkan sampai tidak ada anggota *cluster* yang berpindah *cluster*.

D. Tahap *Performance*

Tahap *performance* adalah tahap pengujian *dataset* dengan menggunakan *tools Rapid Miner* versi 5.3.000. Adapun hasil dari pengujian berupa banyaknya *cluster* yang dibentuk, jumlah anggota pada masing-masing *cluster*, inisial anggota *cluster*, penyebaran anggota *cluster*, dan nilai *centroid* akhir.

E. Tahap Analisis

Pada tahap analisis hasil dari perhitungan dengan menggunakan *Microsoft Excel 2013* dan *tools Rapid Miner* versi 5.3.000 dianalisis apakah hasil perhitungan *manual* sama dengan pengujian menggunakan *tools* atau tidak.

IV. HASIL DAN PEMBAHASAN

Hasil perhitungan *manual* dari penelitian ini menggunakan *Microsoft Excel 2013*. Pada perhitungan *manual* ini iterasi (perulangan) yang terjadi adalah sebanyak 10 kali iterasi hingga tidak ada anggota *cluster* yang berpindah ke *cluster* yang lain. Tabel 4 adalah hasil akhir perhitungan *manual*.

Tabel 4 Hasil akhir perhitungan *manual*

Provinsi	Tingkat			Centroid			Jarak
	VII	VIII	IX	C1	C2	C3	Terpendek
1	0.29	0.33	0.81	<b>0.19</b>	0.34	2.30	0.19
2	0.27	0.32	0.99	0.36	<b>0.18</b>	2.12	0.18
3	0.37	0.37	1.16	0.55	<b>0.07</b>	1.95	0.07

Provinsi	Tingkat			Centroid			Jarak
	VII	VIII	IX	C1	C2	C3	Terpendek
4	0.31	0.31	0.61	<b>0.06</b>	0.54	2.50	0.06
5	0.17	0.33	0.48	<b>0.18</b>	0.68	2.63	0.18
6	0.31	0.33	1.16	0.53	<b>0.09</b>	1.95	0.09
7	0.33	0.27	0.76	<b>0.15</b>	0.41	2.35	0.15
8	0.23	0.31	1.14	0.51	<b>0.13</b>	1.98	0.13
9	0.52	0.32	0.38	<b>0.36</b>	0.79	2.73	0.36
10	0.32	0.34	0.68	<b>0.10</b>	0.47	2.43	0.10
11	0.21	0.19	0.45	<b>0.21</b>	0.73	2.68	0.21
12	0.33	0.33	0.80	<b>0.19</b>	0.35	2.31	0.19
13	0.27	0.31	1.02	0.39	<b>0.16</b>	2.09	0.16
14	0.55	0.57	0.89	0.48	<b>0.38</b>	2.22	0.38
15	0.29	0.25	0.71	<b>0.09</b>	0.46	2.41	0.09
16	0.33	0.38	0.99	0.38	<b>0.16</b>	2.12	0.16
17	0.34	0.43	0.78	<b>0.22</b>	0.36	2.32	0.22
18	0.30	0.40	0.85	<b>0.25</b>	0.29	2.25	0.25
19	0.23	0.28	0.62	<b>0.03</b>	0.54	2.50	0.03
20	0.21	0.29	0.66	<b>0.06</b>	0.51	2.46	0.06
21	0.21	0.64	0.83	0.41	<b>0.40</b>	2.28	0.40
22	0.09	0.15	0.23	<b>0.46</b>	0.97	2.91	0.46
23	0.41	0.37	0.51	<b>0.21</b>	0.64	2.60	0.21
24	0.26	0.31	0.60	<b>0.04</b>	0.55	2.51	0.04
25	0.23	0.32	0.81	<b>0.18</b>	0.35	2.30	0.18
26	0.16	0.27	1.42	0.79	<b>0.35</b>	1.71	0.35
27	0.24	0.27	0.76	<b>0.13</b>	0.41	2.36	0.13
28	0.17	0.15	0.80	<b>0.23</b>	0.45	2.34	0.23
29	0.16	0.12	0.77	<b>0.23</b>	0.50	2.37	0.23
30	0.07	0.14	0.20	<b>0.49</b>	1.01	2.94	0.49
31	0.32	0.31	1.13	0.50	<b>0.10</b>	1.98	0.10
32	0.32	0.49	1.45	0.85	<b>0.32</b>	1.65	0.32
33	0.35	0.53	3.10	2.48	1.96	<b>0.00</b>	0.00
34	0.41	0.66	1.51	0.97	<b>0.45</b>	1.60	0.45

Dapat dilihat pada Tabel 4 bahwa anggota *cluster* 1 sebanyak 21 anggota, *cluster* 2 sebanyak 12 anggota dan *cluster* 3 sebanyak 1 anggota. Adapun pusat *cluster* yang baru pada iterasi 10 seperti pada Tabel 5.

Tabel 5 Pusat *cluster* iterasi 10

Pusat Cluster	Tingkat		
	VII	VIII	IX
C1	0.261	0.281	0.632
C2	0.312	0.413	1.141
C3	0.350	0.530	3.100

Dari Tabel 5 diketahui bahwa iterasi terjadi sebanyak 10 kali iterasi sampai pusat *cluster* pada iterasi 9 sama dengan iterasi 10. Sehingga proses *K-Means Clustering* dihentikan pada iterasi 10. Pusat *cluster* pada iterasi 10 seperti pada Tabel 5 adalah hasil dari pengujian menggunakan *tools Rapid Miner* versi 5.3.000, sehingga menggunakan tiga angka dibelakang titik, sedangkan pada perhitungan *manual* menggunakan pembulatan dua angka di belakang titik. Adapun inisial anggota dari masing-masing *cluster* disajikan pada Tabel 6.

Tabel 6 Inisial Anggota masing-masing *cluster*

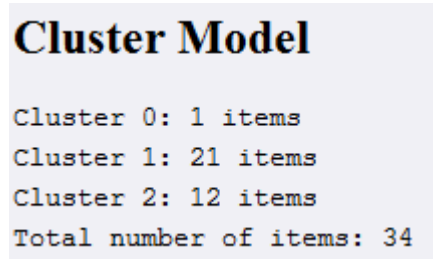
No.	Cluster	Anggota	Jumlah
1	C1 (Rendah)	1,4,5,7,9,10,11,12,15,17,18,19,20,22,23,24,25,27,28,29,30	21
2	C2 (Sedang)	2,3,6,8,13,14,16,21,26,31,32,34	12
3	C3 (Tinggi)	33	1

Inisial pada Tabel 6 kemudian dijabarkan menjadi nama provinsi kembali seperti tertampil dalam tabel 7.

Tabel 7 Provinsi anggota masing-masing *cluster*

No	Cluster	Anggota	Jumlah
1	C1 (Rendah)	DKI Jakarta, Jawa Tengah, DI Yogyakarta, Aceh, Sumatera Barat, Riau, Kepulauan Riau, Jambi, Bengkulu, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Gorontalo, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Maluku, Maluku Utara, Bali	21
2	C2 (Sedang)	Jawa Barat, Banten, Jawa Timur, Sumatera Utara, Sumatera Selatan, Bangka Belitung, Lampung, Kalimantan Utara, Sulawesi Barat, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua Barat	12
3	C3 (Tinggi)	Papua	1

Berikut hasil pengujian menggunakan *tools Rapid Miner* versi 5.3.000 seperti ditunjukkan pada Gambar 2.



Gambar 2 Jumlah anggota cluster

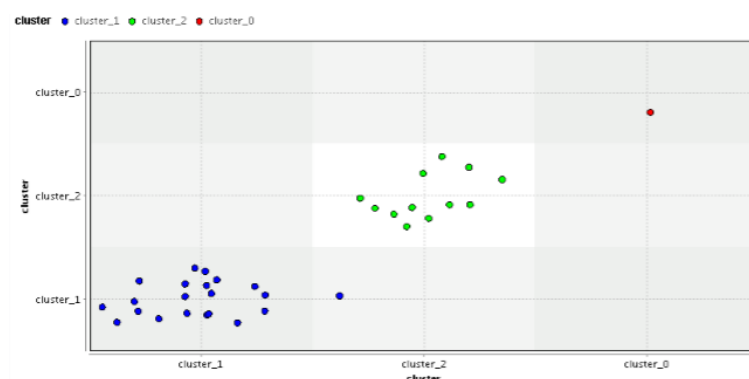
Seperti pada Gambar 2, hasil dari pengujian menggunakan *tools Rapid Miner* versi 5.3.000 sama seperti hasil perhitungan *manual* menggunakan *Microsoft Excel 2013* seperti pada Tabel 4 yaitu cluster 0 pada *tools* sma dengan cluster 3 pada *Ms. Excel* dengan 1 anggota, cluster 1 pada *tools* sma dengan cluster 1 pada *Ms. Excel* dengan 21 anggota, dan cluster 2 pada *tools* sma dengan cluster 2 pada *Ms. Excel* dengan 12 anggota. Selanjutnya adalah hasil inisial anggota cluster pada pengujian menggunakan *tools Rapid Miner* versi 5.3.000 seperti pada Gambar 3.

Row No.	id	cluster	VII	VIII	IX
1	1	cluster_1	0.290	0.330	0.810
2	2	cluster_2	0.270	0.320	0.990
3	3	cluster_2	0.370	0.370	1.160
4	4	cluster_1	0.310	0.310	0.610
5	5	cluster_1	0.170	0.330	0.480
6	6	cluster_2	0.310	0.330	1.160
7	7	cluster_1	0.330	0.270	0.760
8	8	cluster_2	0.230	0.310	1.140
9	9	cluster_1	0.520	0.320	0.380
10	10	cluster_1	0.320	0.340	0.680
11	11	cluster_1	0.210	0.190	0.450
12	12	cluster_1	0.330	0.330	0.800
13	13	cluster_2	0.270	0.310	1.020
14	14	cluster_2	0.550	0.570	0.890
15	15	cluster_1	0.290	0.250	0.710
16	16	cluster_2	0.330	0.380	0.990
17	17	cluster_1	0.340	0.430	0.780
18	18	cluster_1	0.300	0.400	0.850
19	19	cluster_1	0.230	0.280	0.620
20	20	cluster_1	0.210	0.290	0.660
21	21	cluster_2	0.210	0.640	0.830
22	22	cluster_1	0.090	0.150	0.230
23	23	cluster_1	0.410	0.370	0.510
24	24	cluster_1	0.260	0.310	0.600
25	25	cluster_1	0.230	0.320	0.810
26	26	cluster_2	0.160	0.270	1.420
27	27	cluster_1	0.240	0.270	0.760
28	28	cluster_1	0.170	0.150	0.800
29	29	cluster_1	0.160	0.120	0.770
30	30	cluster_1	0.070	0.140	0.200
31	31	cluster_2	0.320	0.310	1.130
32	32	cluster_2	0.320	0.490	1.450
33	33	cluster_0	0.350	0.530	3.100
34	34	cluster_2	0.410	0.660	1.510

Gambar 3 Inisial anggota cluster

Seperti pada Gambar 3, hasil dari pengujian menggunakan *tools Rapid Miner* versi 5.3.000 sama seperti hasil perhitungan *manual* menggunakan *Microsoft Excel 2013* seperti pada Tabel 6 yaitu cluster

0 pada *tools* sama dengan *cluster* 3 pada *Ms. Excel*, *cluster* 1 pada *tools* sama dengan *cluster* 1 pada *Ms. Excel*, dan *cluster* 2 pada *tools* sama dengan *cluster* 2 pada *Ms. Excel*. Selanjutnya adalah hasil penyebaran anggota *cluster* pada pengujian menggunakan *tools Rapid Miner* versi 5.3.000 seperti pada Gambar 4.



Gambar 4 Penyebaran anggota *cluster*

Pada Gambar 4 titik warna biru menunjukkan penyebaran anggota *cluster* pada *cluster* 1 dengan banyaknya titik menunjukkan jumlah anggota *cluster* yaitu 21 anggota. Warna hijau menunjukkan penyebaran anggota *cluster* 2 dengan banyaknya anggota adalah 12. Sedangkan titik merah pada Gambar 4 menunjukkan penyebaran anggota *cluster* 3 yaitu sebanyak 1 anggota. Dapat dikatakan bahwa perhitungan secara *manual* dan pengujian menggunakan *tools* hasilnya adalah sama. Hal ini dapat dilihat dari hasil akhir pengujian yaitu berupa banyaknya anggota *cluster* pada *cluster* 1, *cluster* 2, dan *cluster* 3 serta inisial anggota yang ada pada masing-masing *cluster* tersebut. Selain kedua hal tersebut hasil nilai pusat *cluster* (*centroid*) antara perhitungan *manual* maupun menggunakan *tools* sama.

## V. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilaksanakan maka dapat ditarik kesimpulan bahwa penelitian ini telah berhasil menerapkan penggunaan algoritma *K-Means* untuk mengelompokkan tingkat putus sekolah jenjang SMP tahun ajar 2017/2018 di Indonesia. *Cluster* yang dihasilkan adalah 3 *cluster* yaitu C1 menunjukkan tingkat putus sekolah jenjang SMP rendah, C2 menunjukkan tingkat putus sekolah jenjang SMP sedang, dan C3 menunjukkan tingkat putus sekolah jenjang SMP tinggi. Jumlah anggota *cluster* menurut perhitungan secara *manual* menggunakan *Microsoft Excel 2013* maupun menggunakan *tools Rapid Miner* versi 5.3.000 menunjukkan hasil yang sama yaitu jumlah anggota pada *cluster* 1 sebanyak 21 anggota, *cluster* 2 sebanyak 12 anggota, dan *cluster* 3 sebanyak 1 anggota.

Dari hasil perhitungan *manual* dan *tools* menunjukkan nilai *centroid* akhir sama yaitu *cluster* 1 dengan nilai *centroid* akhir untuk tingkat VII sebesar 0.261, tingkat VIII sebesar 0.281, dan tingkat IX sebesar 0.632. Selanjutnya nilai *centroid* akhir pada *cluster* 2 untuk tingkat VII sebesar 0.312, tingkat VIII sebesar 0.413, dan tingkat IX sebesar 1.141. Dan nilai *centroid* akhir pada *cluster* 3 untuk tingkat VII sebesar 0.350, tingkat VIII sebesar 0.530, dan tingkat IX sebesar 3.100 dengan catatan perhitungan *manual* dilakukan pembulatan dua angka dibelakang titik.

Berdasarkan hasil pengelompokan provinsi di Indonesia dengan tingkat putus sekolah jenjang SMP rendah, sedang dan tinggi, diharapkan ada tindak lanjut dari pemerintah daerah dan pusat agar dapat mengevaluasi angka putus sekolah yang terjadi di Indonesia dan dapat mengambil kebijakan agar angka putus sekolah jenjang SMP di Indonesia dapat dikurangi pada tahun ajar selanjutnya.

## PENGAKUAN

Naskah Ilmiah ini adalah sebagian dari penelitian Tugas Akhir milik Sekar Wuni dengan judul Implementasi Algoritma *K-Means* untuk Mengelompokkan Tingkat Putus Sekolah Jenjang SMP di Indonesia yang dibimbing oleh Amril Mutoi Siregar dan Dwi Sulistya Kusumaningrum.

## DAFTAR PUSTAKA

- [1] Indonesia. (1945). UUD RI Tahun 1945, (2), 1–19. <https://doi.org/10.1007/s13398-014-0173-7.2>
- [2] Noor Rizqa. (2015). Skripsi Faktor Penyebab Anak Putus Sekolah Pada Tingkat Smp Di Desa Bumi Rejo Kecamatan Baradatu Kabupaten Way Kanan Tahun 2014. *Biomass Chem Eng*, 49(23–6).

- [3] Kemendikbud. (2017). Ikhtisar Data Pendidikan dan Kebudayaan Tahun 2017/2018, 1–61. Retrieved from <http://publikasi.data.kemdikbud.go.id>
- [4] Kemendikbud. (2018). *Statistik Persekolahan SMP 2017/2018* (1st ed.). Jakarta Pusat: PDSPK Kemendikbud. Retrieved from <http://repositori.kemdikbud.go.id>
- [5] Pang-Ning Tan, Michael Steinbach, V. K. (2006). Introduction to *Data mining*. [https://doi.org/10.1007/430\\_2009\\_1](https://doi.org/10.1007/430_2009_1)
- [6] Jiawei Han, Micheline Kamber, J. P. (2012). *Data mining – Concepts & Techniques*. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- [7] Aditya, K. B., Setiawan, Y., & Puspitaningrum, D. (2018). Sistem Informasi Geografis Pemetaan Faktor-Faktor Yang Mempengaruhi Angka Kematian Ibu (Aki) Dan Angka Kematian Bayi (Akb) Dengan Metode K-Means Clustering (Studi Kasus: Provinsi Bengkulu). *Jurnal Teknik Informatika*, 10(1), 59–66. <https://doi.org/10.15408/jti.v10i1.6817>
- [8] Windarto, A. P. (2017). Implementation of *Data mining* on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method. *International Journal of Artificial Intelligence Research*, 1(2), 26. <https://doi.org/10.29099/ijair.v1i2.17>
- [9] Suryanto, J. (2017). Analisa Perbandingan Pengelompokkan Curah Hujan 15 Harian Provinsi DIY Menggunakan Fuzzy Clustering Dan K-Means Clustering, XVI, 229–242
- [10] Siregar, A. M., & Puspabhuana, A. (2017). *Data mining*, I (ISBN: 978-602-5473-96-8), 1–198.
- [11] Prasetyo, E. (2014). *Data mining Mengolah Data Menjadi Informasi menggunakan Matlab*. (A. Sahala, Ed.) (1st ed.). Yogyakarta: Penerbit Andi.
- [12] Prasetyo, E. (2012). *Data mining Konsep dan Aplikasi menggunakan Matlab*. (N. WK, Ed.) (1st ed.). Yogyakarta: Penerbit Andi.
- [13] Larose, D. T. (2005). *Discovering Knowledge In Data An Introduction to Data mining*. <https://doi.org/http://dx.doi.org/10.18637/jss.v016.b01>
- [14] Gudono. (2011). Analisis Data Multivariat.
- [15] Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On Clustering Validation Techniques, 107–145.
- [16] Suprihatin. (2011). Klastering K-means untuk Penentuan Nilai Ujian, 1(1), 53–62.
- [17] Musthafawi, A. Z., Rosiani, U. D., & Atmoko, A. (2017). Analisis Respon Emosi Marah Wanita Jawa Dengan Algoritma K-Means Clustering, 33–38.
- [18] Zubair Khan , Jianjun Ni , Xinnan Fan, and P. S. (2017). an Improved K -Means Clustering Algorithm Based on an Adaptive Initial Parameter Estimation Procedure, 13(5).
- [19] Zulfadhilah, M. (2016). Log Classification using K-Means Clustering for Identify Internet User Behaviors. *Internatonal Journal of Computer Application*, 154(3), 34–39. <https://doi.org/10.5120/ijca2016912076>
- [20] Aranda, J., & Natasya, W. A. G. (2016). Penerapan Metode K-Means Cluster Analysis Pada Sistem Pendukung Keputusan Pemilihan Konsentrasi Untuk Mahasiswa International Class Stmik Amikom Yogyakarta. *Semnasteknomedia Online*, 4(1), 4-2–1. Retrieved from <https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/view/129>
- [21] Herawatie, D. (2014). Perbandingan Algoritma Pengelompokan Non-Hierarki untuk Dataset Dokumen. Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta, 11–16.
- [22] Soni, N., & Ganatra, A. (2012). Categorization of Several Clustering Algorithms from Different Perspective: A Review, 2(8), 63–68.
- [23] Maulida, L. (2018). Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov . Dki Jakarta Dengan K-Means. *JISKa*, 2(3), 167–174.