

PENGELOMPOKAN DATA SISWA DI INDONESIA MENGGUNAKAN K-MEANS CLUSTERING

STUDENT DATA GROUPING IN INDONESIA USING K-MEANS CLUSTERING

Rizky Maulana, Fairuz Azmi, ST., MT.², Dr. Purba Daru Kusuma S.T., M.T.³

^{1,2,3}Prodi S1 Teknik Komputer, Fakultas Teknik Elektro, Universitas Telkom

¹lonewolf@student.telkomuniversity.ac.id, ²worldliner@telkomuniversity.co.id, ³purbodaru@gmail.ac.id

Abstrak

Tingginya angka siswa putus sekolah di Indonesia menjadi kasus yang tidak pernah lepas dari perhatian pemerintah. Akibat kurangnya pembangunan mutu pendidikan, faktor ekonomi, dan susah akses ke sekolah, secara kuantitas muncul jumlah angka anak putus sekolah yang dominan pada provinsi tertentu. Dari permasalahan di atas, dapat dianalisis jumlah siswa putus sekolah di Indonesia dengan pengelompokan jumlah siswa putus sekolah di setiap provinsi serta jumlah sekolah dan jumlah siswa di setiap provinsi di Indonesia. Pada tugas akhir ini dirancang program aplikasi berbasis web untuk mengelompokkan data siswa di Indonesia. Metode yang digunakan untuk mengelompokkan data tersebut adalah K-Means Clustering. Keluaran dari tugas akhir ini adalah analisis hasil dari pengelompokan data siswa pada setiap provinsi di Indonesia dan pengujian stabilitas dari K-Means Clustering. Dari hasil pengujian stabilitas clustering, didapatkan standar deviasi terendah yaitu 0 dan yang tertinggi yaitu 4.85 Hasil penelitian ini dapat membantu Dinas Pendidikan dalam mengatasi masalah siswa putus sekolah di Indonesia pada setiap provinsi di Indonesia.

Kata Kunci: Pendidikan, data, K-Means Clustering, Standar Deviasi

Abstract

The high number of students dropping out of school in Indonesia is a case that has never been separated from the government's attention. As a result of the lack of development in the quality of education, economic factors, and the difficulty of access to schools, in terms of quantity the number of school dropouts appears dominant in certain provinces. From the above problems, it can be analyzed the number of dropout students in Indonesia by grouping the number of dropout students in each province as well as the number of schools and the number of students in each province in Indonesia. In this final project a web-based application program is designed to classify student data in Indonesia. The method used to classify the data is K-Means Clustering. The output of this thesis is an analysis of the results of grouping student data in each province in Indonesia and testing the stability of K-Means Clustering. From the results of clustering stability testing, obtained the lowest standard deviation is 0 and the highest is 4.85 The results of this study can help the Office of Education in overcoming the problem of dropout students in Indonesia in every province in Indonesia.

Keywords: Education, data, K-Means Clustering, Standard Deviation

1. Pendahuluan

Pendidikan adalah suatu hak dan kewajiban bagi seluruh warga negara Indonesia seperti yang tercantum dalam UUD 1945 Pasal 31 Ayat 1 dan 2, setiap warga negara berhak mendapatkan pendidikan dan warga negara wajib mengikuti pendidikan dasar [1]. Dengan kualitas pendidikan yang baik dan merata sudah halnya menjadi hak warga negara Indonesia. Untuk itu, sudah seharusnya menjadi kewajiban pemerintah untuk menyediakan sarana dan prasarana yang menunjang pendidikan.

Salah satu masalah pendidikan yang terjadi di Indonesia yaitu tingginya angka putus sekolah, terutama pada daerah padat penduduk dan daerah perdesaan (terpencil). Salah satu pengamat pendidikan dan penggagas berbagai gerakan di bidang pendidikan, Najeela Shihab mengungkapkan, salah satu permasalahan pendidikan utama di Indonesia yaitu susah akses ke sekolah, sehingga menyebabkan anak putus sekolah [2]. Sementara itu, Abduh Zen, Ketua Litbang PB PGRI dan Direktur Institute for Education Reform menyatakan faktor ekonomi dan kemiskinan yang menjadi penyebab terbesar anak putus sekolah [3]. Berdasarkan data UNESCO tahun 2016 anak putus sekolah di Indonesia diperkirakan 1 juta anak per tahunnya [4]. Dan pada tahun 2017, Kementerian Pendidikan dan Kebudayaan mencatat lebih dari 4.1 juta anak berusia 6-21 tahun tidak sekolah [5].

Diperlukan proses analisa mengenai siswa putus sekolah di Indonesia. Selain itu, perkembangan pada bidang teknologi komputasi memungkinkan analisa data secara otomatis. Teknik *clustering* adalah salah satu teknik komputasi yang umum untuk analisa data. Adapun salah satu metode *clustering* yang cukup populer adalah *K-Means Clustering*. *K-Means Clustering* telah digunakan untuk analisa data di berbagai bidang. Oleh Karena itu, analisa data kondisi siswa di Indonesia menggunakan metode *K-Means Clustering* dapat dimungkinkan.

2. Dasar Teori

2.1 Pengertian Siswa

Pengertian siswa dalam Kamus Besar Bahasa Indonesia adalah murid/pelajar/anak (orang yang sedang berguru/belajar, bersekolah terutama pada tingkat sekolah dasar dan menengah). Sedangkan menurut UU No.20

Tahun 2003 tentang Pendidikan Nasional, menyatakan bahwa siswa adalah “adalah anggota masyarakat yang berusaha mengembangkan potensi diri melalui proses pembelajaran yang tersedia pada jalur, jenjang, dan jenis pendidikan tertentu”. Definisi siswa secara umum di Indonesia adalah peserta didik yang menempuh pendidikan dibangku SD (Sekolah Dasar), SMP (Sekolah Menengah Pertama), SMA (Sekolah Menengah Atas), SMK (Sekolah Menengah Kejuruan). Siswa atau peserta didik mengikuti pembelajaran di sekolah untuk menjadi individu yang mampu bersaing, beretika, dan sopan santun dengan masyarakat.

2.2 Clustering

Clustering adalah metode penganalisaan data yang sering dimasukan sebagai salah satu metode data *mining* yang bertujuan untuk mengelompokkan data dengan karakteristik yang sama ke suatu 'wilayah' yang sama dan data dengan karakteristik yang berbeda ke 'wilayah' yang lain [6]. Tujuan dari proses clustering yaitu untuk mengelompokkan data ke dalam suatu klaster, sehingga objek pada suatu klaster memiliki kemiripan yang sangat besar dengan objek lain pada klaster yang sama [7].

Ada beberapa karakteristik dari clustering, yaitu [8]:

1. Partitioning clustering

- Disebut juga exclusive clustering
- Setiap data harus termasuk kedalam cluster tertentu
- Mengungkinkan bagi setiap data yang termasuk cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke cluster yang lain.

Contoh: K-Means dan Residual Analysis.

2. Hierarchical clustering

- Setiap data harus masuk kedalam cluster tertentu
- Suatu data termasuk ke dalam cluster tertentu pada suatu proses, tidak dapat berpindah ke cluster lain.

Contoh: Single Linkage, Centroid Linkage, Complete Linkage, Average Linkage.

3. Overlapping Clustering

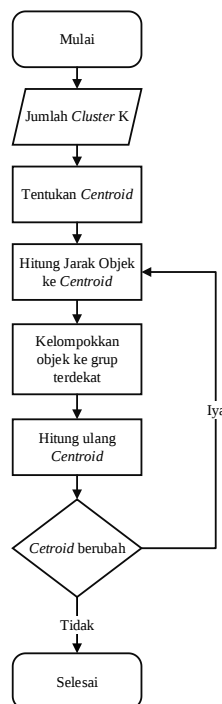
- Setiap data memungkinkan termasuk ke beberapa cluster
- Data mempunyai nilai keanggotaan pada beberapa cluster

Contoh : Fuzzy C-means, Gaussian Mixture.

4. Hybrid Merupakan kombinasi dari karakteristik partitioning, overlapping dan hierarchical

2.3 K-Means Clustering

K-Means adalah metode clustering *non hierarchical* berbasis jarak yang membagi data ke dalam cluster dan algoritma ini bekerja pada atribut numerik. *K-means clustering* sebagai salah satu metode data clustering non-hirarki mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain [9].



Gambar 2.1 Flowchart K-Means Clustering

Algoritma *K-Means Clustering* dimulai dengan pemilihan K buah titik pusat (*centroid*) secara acak. Setelah itu, hitung jarak data ke semua centroid dan kelompokkan data tersebut berdasarkan jarak terdekat ke centroid, hitung rata-rata nilai data setiap kelompok perbaharui nilai titik *centroid*. Lakukan pengulangan hingga nilai dari titik *centroid* tidak lagi berubah. Perhitungan jarak data ke- i (x_i) pada pusat cluster ke- k (c_k) diberi nama (d_{ik}), dapat menggunakan rumus Euclidean, yaitu [10] :

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

Keterangan :

$d(x, y)$ = jarak titik x dengan titik y

x = nilai x ($x_1, x_2, \dots, x_n$)

y = nilai y ($y_1, y_2, \dots, y_n$)

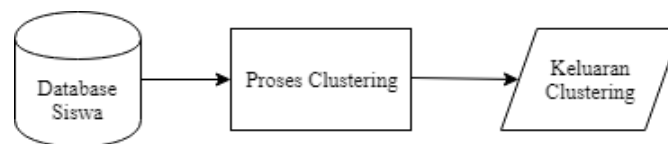
n = jumlah variabel

i = jumlah data

3. Pembahasan

3.1. Gambaran Umum Sistem

Pada tahap penelitian ini, algoritma *K-Means Clustering* diterapkan ke dalam aplikasi pengelompokan berbasis *web*. Aplikasi ini dikembangkan dengan menggunakan Bahasa pemrograman *PHP* dan data disimpan ke dalam system berbasis data *MySQL* sebagai sumber penyimpanannya. Proses pengelompokan dilihat pada gambar 3.1. Data yang digunakan dalam analisa penelitian ini adalah data kementrian pendidikan dan budaya tahun ajaran 2017/2018. Data tersebut meliputi data jumlah siswa putus sekolah, jumlah siswa, dan jumlah sekolah setiap provinsinya. Data pengelompokan yang digunakan ini dibagi ke beberapa tingkat jenjang pendidikan mulai dari SD, SMP, SMA, dan SMK.



Gambar 3.1 Proses clustering

Data yang sudah ada kemudian di analisa ke dalam beberapa pengelompokan yang terdiri dari : pengelompokan jumlah siswa putus sekolah, pengelompokan jumlah siswa, pengelompokan perbandingan jumlah siswa dan siswa putus sekolah, pengelompokan perbandingan jumlah sekolah dan siswa putus sekolah. Pengelompokan akan didasarkan pada tiap tingkat jenjang pendidikan.

3.2. Basis Data

Pada penelitian ini, desain basis data yang digunakan adalah *MySQL* yang merupakan implementasi dari *Relational Database Management System* (RDBMS) dengan menggunakan bahasa pemrograman *PHP* sebagai kontrolnya. Data disimpan ke dalam software open source *phpMyAdmin*. Pada software tersebut data dimasukkan ke dalam database *cluster_putus_sekolah* yang memiliki 4 tabel data yaitu : *tbl_provinsi*, *tbl_putus*, *tbl_sekolah*, dan *tbl_siswa*.

Tabel *tbl_provinsi* berisikan data mengenai provinsi yang ada di Indonesia. Isi dari data yang ada pada tabel tersebut meliputi : *id_prov* dengan tipe integer dan nama dengan tipe *varchar*.

Tabel *tbl_putus* berisikan data mengenai jumlah siswa putus sekolah pada setiap provinsi di Indonesia. Isi dari data yang ada pada tabel tersebut meliputi : *provinsi* dengan tipe integer, *putus_sd* dengan tipe integer, *putus_smp* dengan tipe integer, *putus_sma* dengan tipe integer, *putus_smk* dengan tipe integer.

Tabel *tbl_sekolah* berisikan data mengenai jumlah sekolah pada setiap provinsi di Indonesia. Isi dari data yang ada pada tabel tersebut meliputi : *provinsi* dengan tipe integer, *sekolah_sd* dengan tipe integer, *sekolah_smp* dengan tipe integer, *sekolah_sma* dengan tipe integer, *sekolah_smk* dengan tipe integer.

Tabel *tbl_siswa* berisikan data mengenai jumlah siswa pada setiap provinsi di Indonesia. Isi dari data yang ada pada tabel tersebut meliputi : *provinsi* dengan tipe integer, *siswa_sd* dengan tipe integer, *siswa_smp* dengan tipe integer, *siswa_sma* dengan tipe integer, *siswa_smk* dengan tipe integer.

4. Analisis

Pada bagian ini, dibahas hasil pengelompokan data siswa di Indonesia. Analisis ini berfokus pada distribusi dan perbedaan data siswa di setiap clusternya. Parameter dari *output* yang akan diamati di setiap cluster meliputi : Jumlah provinsi, jumlah rata-rata, jumlah minimal, jumlah maksimum, dan list provinsi. Hasil pengelompokan yang dianalisa merupakan data keseluruhan dari tingkat jenjang pendidikan mulai dari SD, SMP, SMA, dan SMK yang meliputi data jumlah siswa, jumlah siswa putus sekolah, perbandingan siswa putus sekolah dan jumlah siswa, dan perbandingan siswa putus sekolah dan jumlah sekolah.

Tabel 4.1 Hasil Clustering Data Siswa Putus Sekolah Di Indonesia

No. Cluster	Total siswa putus sekolah	Jumlah Anggota	Rata - Rata Siswa Putus Sekolah	Minimal Siswa Putus Sekolah	Maksimal Siswa Putus Sekolah	List Provinsi
1	9680	9	1075.555556	528	1411	Kepulauan Riau, Bangka Belitung, Kalimantan Utara, Sulawesi Utara, Gorontalo, Sulawesi Barat, Maluku, Bali, Papua Barat,
2	17509	9	1945.444444	1558	2510	DI Yogyakarta, Jambi, Bengkulu, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Tengah, Sulawesi Tenggara, Maluku Utara,
3	23966	6	3994.333333	3480	4919	Aceh, Sumatera Barat, Riau, Kalimantan Barat, Nusa Tenggara Barat, Papua,
4	42014	6	7002.333333	6235	8466	DKI Jakarta, Banten, Sumatera Selatan, Lampung, Sulawesi Selatan, Nusa Tenggara Timur,
5	94659	4	23664.75	15995	35209	Jawa Barat, Jawa Tengah, Jawa Timur, Sumatera Utara,

Pada Tabel 4.1 menampilkan hasil dari pengelompokan data siswa putus sekolah di Indonesia yang menunjukkan jumlah anggota provinsi pada setiap clusternya dengan distribusi yang cenderung menurun. Pada cluster pertama dan kedua, rata-rata siswa putus sekolah di Indonesia memiliki nilai terendah dibandingkan dengan cluster lain serta dengan jumlah provinsi terbanyak. Berbeda dengan cluster kelima yang memiliki rata-rata jumlah siswa putus sekolah di Indonesia terbanyak namun dengan jumlah provinsi yang paling sedikit. Pada rata-rata siswa putus sekolah mengalami kenaikan yang bervariasi setiap clusternya. Sehingga membuat perbedaan yang sangat signifikan antara cluster pertama dan cluster kelima yaitu kurang lebih 2 kali lipat. Dari total jumlah siswa putus sekolah di setiap cluster, cluster kelima memiliki total jumlah siswa putus sekolah terbanyak walau hanya dengan 4 provinsi saja.

Dari aspek persebaran provinsi, provinsi di pulau Jawa tersebar pada cluster kedua, keempat, dan kelima dengan tiga provinsi yang memiliki jumlah rata-rata siswa putus sekolah terbanyak dari provinsi lain yaitu Jawa Barat, Jawa Tengah, dan Jawa Timur. Selain itu, provinsi-provinsi diluar pulau Jawa khususnya provinsi yang berada di timur Indonesia seperti Papua, Papua Barat, Maluku, Maluku Utara tersebar pada cluster pertama sampai ketiga dengan jumlah siswa putus sekolah yang terbilang rendah.

Tabel 4.2 Hasil Clustering Data Jumlah Siswa Di Indonesia

No. Cluster	Total Siswa	Jumlah Anggota	Rata - Rata Siswa	Minimal Siswa	Maksimal Siswa	List Provinsi
1	2586771	9	287419	135380	417175	Kepulauan Riau, Bangka Belitung, Bengkulu, Kalimantan Utara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat,
2	5339037	9	593226.33	470569	716369	DI Yogyakarta, Jambi, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Papua,
3	5935418	6	989236.33	777781	1305175	Aceh, Sumatera Barat, Riau, Kalimantan Barat, Bali, Nusa Tenggara Barat,
4	12815188	7	1830741.14	1374055	3069381	DKI Jakarta, Banten, Sumatera Utara, Sumatera Selatan, Lampung, Sulawesi Selatan, Nusa Tenggara Timur,
5	18623492	3	6207830.67	5231156	8026455	Jawa Barat, Jawa Tengah, Jawa Timur,

Pada Tabel 4.2 menampilkan hasil pengelompokan data jumlah siswa di Indonesia yang menunjukkan jumlah provinsi disetiap clusternya terdistribusi miring positif. Jumlah rata-rata siswa di Indonesia pada cluster pertama hingga keempat mengalami kenaikan kurang lebih 2 kali lipat dari cluster sebelumnya. Namun pada cluster kelima mengalami kenaikan hingga lebih dari 3 kali lipat dari cluster keempat yang mengakibatkan perbedaan yang sangat signifikan antara cluster pertama dan kelima dengan perbedaan lebih dari 22 kali lipat. Pada total jumlah siswa setiap cluster mengalami kenaikan yang bervariasi dengan total jumlah siswa terbanyak pada cluster kelima dengan 3 provinsi.

Provinsi-provinsi di pulau Jawa tersebar pada cluster kedua, keempat, dan kelima serta beberapa provinsi diluar pulau Jawa yang berada pada cluster keempat seperti Sumatera Utara, Sumatera Selatan, Lampung, Sulawesi Selatan, dan Nusa Tenggara Timur yang memiliki jumlah rata-rata siswa yang tinggi. Berbeda dengan cluster pertama yang memiliki jumlah rata-rata siswa yang rendah dengan didominasi provinsi-provinsi di luar pulau Jawa seperti Kepulauan Riau, Bangka Belitung, Bengkulu, Kalimantan Utara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, dan Papua barat.

Tabel 4.3 Hasil Clustering Perbandingan Jumlah Siswa Dan Siswa Putus Sekolah Di Indonesia

No. Cluster	Jumlah Anggota	Rata - Rata Siswa Putus Sekolah	Minimal Siswa Putus Sekolah	Maksimal Siswa Putus Sekolah	List Provinsi
1	7	184.56	135.81	203.71	Sumatera Utara, Bengkulu, Sulawesi Barat, Maluku Utara, Nusa Tenggara Timur, Papua, Papua Barat,
2	7	226.9	213.97	233.03	Jawa Barat, Jawa Timur, Sumatera Selatan, Bangka Belitung, Lampung, Gorontalo, Nusa Tenggara Barat,
3	8	256.01	242.54	282.66	DKI Jakarta, Banten, Aceh, Jambi, Kalimantan Barat, Kalimantan Tengah, Kalimantan Utara, Sulawesi Selatan,
4	10	318.76	284.64	377.53	Jawa Tengah, DI Yogyakarta, Sumatera Barat, Riau, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Maluku,
5	2	569.78	443.24	696.31	Kepulauan Riau, Bali,

Tabel 4.3 menampilkan hasil pengelompokan rasio perbandingan jumlah siswa dengan siswa putus sekolah setiap provinsi yang ada di Indonesia. Dari tabel tersebut, jumlah provinsi pada setiap cluster memiliki distribusi miring negatif. Pada rata-rata rasio jumlah siswa dan siswa putus sekolah terjadi kenaikan yang bervariasi pada setiap clusternya dengan perbedaan pada cluster kelima kurang lebih 5 kali dari cluster pertama.

Pada aspek persebaran provinsi pada tabel tersebut, setengah dari jumlah provinsi yang ada di Indonesia berada pada cluster pertama dan kedua dengan nilai rasio yang rendah. Provinsi yang berada di pulau Jawa tersebar pada cluster kedua sampai keempat. Rasio jumlah siswa dan siswa putus sekolah tertinggi terdapat pada cluster kelima yang hanya di tempati oleh 2 provinsi di luar pulau Jawa yaitu Kepulauan Riau dan Bali. Berbeda dengan cluster pertama dan kedua yang didominasi oleh provinsi di luar pulau Jawa, pada cluster tersebut rasio perbandingan jumlah siswa dan siswa putus sekolah cukup rendah.

Tabel 4.4 Hasil Clustering Perbandingan Jumlah Sekolah dan Siswa Putus Sekolah

No. Cluster	Jumlah Anggota	Rata - Rata Siswa Putus Sekolah	Minimal Siswa Putus Sekolah	Maksimal Siswa Putus Sekolah	List Provinsi
1	2	0.72	0.71	0.73	DKI Jakarta, Papua,
2	5	0.9	0.82	0.96	Jawa Barat, Banten, Sumatera Utara, Sumatera Selatan, Bangka Belitung,
3	8	1.16	1.06	1.3	Jawa Timur, Riau, Bengkulu, Lampung, Kalimantan Timur, Nusa Tenggara Barat, Nusa Tenggara Timur, Papua Barat,
4	13	1.48	1.34	1.77	Jawa Tengah, DI Yogyakarta, Aceh, Sumatera Barat, Kepulauan Riau, Jambi, Kalimantan Barat, Kalimantan Utara, Gorontalo, Sulawesi Selatan, Sulawesi Barat, Sulawesi Tenggara, Maluku Utara,
5	6	2.4	2.11	2.85	Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Maluku, Bali,

Tabel 4.4 menampilkan hasil pengelompokan rasio perbandingan jumlah sekolah dengan siswa putus sekolah yang ada di Indonesia. Dapat dilihat, persebaran jumlah provinsi lebih terfokus pada cluster keempat dan kelima yang memiliki jumlah provinsi lebih dari setengah dari provinsi yang ada di Indonesia. Pada rata-rata rasio siswa putus sekolah, perbedaan di setiap clusternya mengalami kenaikan secara linear. Namun, perbedaan antara rasio pada cluster pertama dan kelima cukup signifikan yaitu lebih dari 3 kali lipat.

Pada pengelompokan ini, provinsi yang ada di pulau Jawa tersebar hampir pada semua cluster kecuali pada cluster kelima dengan beberapa provinsi berada pada cluster yang memiliki rata-rata rasio yang tinggi seperti Jawa Tengah, DI Yogyakarta dan sebagian lain berada pada cluster dengan rata-rata rasio yang rendah seperti DKI Jakarta, Jawa Barat, dan Banten. Sedangkan provinsi yang berada di luar pulau Jawa tersebar di semua cluster. Namun, untuk provinsi yang berada di timur Indonesia tersebar hampir di semua cluster kecuali cluster kedua dengan provinsi Papua pada cluster pertama yang memiliki nilai rata-rata rasio terendah.

5. Kesimpulan

Dari penelitian yang dilakukan dapat disimpulkan bahwa, berdasarkan hasil analisis pengelompokan perbandingan jumlah siswa dan siswa putus sekolah, terdapat kesenjangan antara kondisi jumlah siswa putus sekolah pada provinsi yang berada di timur Indonesia dan beberapa provinsi di pulau Jawa. Daerah Bali dan Sulawesi Utara berada pada *cluster* dengan rasio yang tinggi pada setiap tingkat jenjang pendidikannya sedangkan daerah dengan rasio yang rendah yaitu, Papua, Papua Barat, dan beberapa provinsi di pulau Jawa. Hasil dari pengelompokan perbandingan jumlah sekolah dan siswa putus sekolah tidak dapat menjadi acuan karena variabel jumlah sekolah tidak berpengaruh pada jumlah siswa putus sekolah. Berdasarkan hasil pengujian stabilitas clustering, pengujian data clustering dengan stabilitas terendah yaitu pada tingkat SMA dengan standar deviasi

tertinggi yaitu 4.85 dan yang terendah yaitu 3.5. Sedangkan hasil pengujian dengan stabilitas tertinggi yaitu pada tingkat SMP dengan standar deviasi tertinggi yaitu 0.98 dan yang terendah yaitu 0. Jumlah siswa putus sekolah di Indonesia masih tergolong cukup tinggi terutama pada provinsi yang berada di timur Indonesia dan beberapa provinsi di pulau Jawa.

Daftar Pustaka:

- [1] R. Indonesia, Undang Undang Dasar 1945 Pasal 31 tentang Pendidikan, Jakarta: Sekretariat Negara, 1945.
- [2] A. N. Kholifah and A. Permatasari, "Masalah Pendidikan di Indonesia, Putus Sekolah hingga Salah Jurusan," Viva.co.id, 15 Agustus 2018. [Online]. Available: <https://www.viva.co.id/gaya-hidup/parenting/1065321-masalah-pendidikan-di-indonesia-putus-sekolah-hingga-salah-jurusan>. [Accessed 27 September 2018].
- [3] A. Nabhani, "Antam Fasilitas Ujian Kesetaraan Kejar Paket - Tekan Angka Putus Sekolah," Neraca, 15 April 2017. [Online]. Available: <http://www.neraca.co.id/article/83766/tekan-angka-putus-sekolah-antam-fasilitas-ujian-kesetaraan-kejar-paket>. [Accessed 2018 September 2018].
- [4] I. Rahmawati, "Tingginya Angka Putus Sekolah di Indonesia," CNN Indonesia, 7 Maret 2017. [Online]. Available: <https://student.cnnindonesia.com/edukasi/20170417145047-445-208082/tingginya-angka-putus-sekolah-di-indonesia>. [Accessed 27 September 2018].
- [5] D. Seftiawan, "4,1 Juta Anak Indonesia Tidak Sekolah," Pikiran Rakyat, 4 Juni 2017. [Online]. Available: <https://www.pikiran-rakyat.com/pendidikan/2017/06/04/41-juta-anak-indonesia-tidak-sekolah-402455>. [Accessed 28 September 2018].
- [6] E. Hadinata, Pengembangan Algoritma Penentuan Titik Awal Dalam Metode Clustering Algoritma Fuzzy C-Means, Medan, Indonesia, 2016.
- [7] P. N. Tan, M. Steinbach and V. Kumar, Introduction to Data Mining, Boston: Pearson Education, 2006.
- [8] J. Han and M. Kamber, Data mining: Concept and Techniques, San Francisco: Morgan Kaufmann, 2001.
- [9] A. N. Khomarudin, Teknik Data Mining : Algoritma K-Means Clustering, Bukittinggi: Komunitas eLearning IlmuKomputer.Com, 2016.
- [10] C. Hardyanto, Peningkatan Kualitas Hasil Clustering Menggunakan Algoritma Hierarchical Agglomerative Clustering Kmeans-Particle Swarm Optimization (Studi Kasus: Segmentasi Pasar Film), Bandung: Perpustakaan UNIKOM, 2017.