

# Penggunaan X-Means Clustering Method untuk Mengelompokkan Potensi Sekolah Menengah Unggul di Kabupaten Banyumas

Aditya Wijayanto <sup>#1</sup>, Rifki Adhitama <sup>#2</sup>

# Prodi Rekayasa Perangkat Lunak, Fakultas Teknologi Industri dan Informatika, Institut Teknologi Telkom Purwokerto

Jln. D. I. Panjaitan No. 128 Purwokerto 53147

<sup>1</sup>aditya.wijayanto@ittelkom-pwt.ac.id

<sup>2</sup>rifki@ittelkom-pwt.ac.id

Accepted on 08-09-2019

## Abstract

Sekolah menengah merupakan salah satu jenjang pendidikan yang diwajibkan pemerintah dalam program wajib belajar Sembilan tahun. Maka dari itu informasi pengelompokan sekolah menengah sangat penting untuk pengembangan maupun revitalisasi dari pendidikan pada jenjang tersebut. Salah satu metode yang dapat digunakan adalah metode X-means clustering. Metode X-Means clustering merupakan salah satu penyempurnaan dari metode K-Means clustering. Metode X-means digunakan pada penelitian ini karena dapat membagi cluster secara optimal dengan menggunakan nilai Bayesian Information Criterion (BIC). Penelitian ini menggunakan 118 data sekolah se-kabupaten Banyumas yang diambil dari pangkalan data kemendikbud. Hasil dari penelitian ini adalah empat buah cluster dengan nilai Davis Bouldin Index sebesar 0.846 dengan kelompok sekolah unggul sebanyak 28 sekolah, kelompok baik sebanyak 26 sekolah, kelompok cukup sebanyak 43 sekolah dan kelompok kurang sebanyak 21 sekolah.

**Keywords:** Clustering, Sekolah Menengah, SMA, SMK, X-Means

## I. INTRODUCTION

Sekolah menengah adalah salah satu jenjang pendidikan formal setelah menempuh jenjang pendidikan dasar menurut Undang-Undang No 20 Tahun 2003. Jenjang pendidikan sekolah menengah ini disiapkan agar siswa mampu melanjutkan ke jenjang pendidikan tinggi atau mampu bekerja. Jenis pendidikan di sekolah menengah antara lain pendidikan umum dan kejuruan. Sekolah menengah yang masuk ke pendidikan umum adalah Sekolah Menengah Atas (SMA) dan yang masuk ke pendidikan kejuruan adalah Sekolah Menengah Kejuruan (SMK). Penyelenggara pendidikan menengah atau sekolah dibedakan menjadi dua, Sekolah negeri dan sekolah swasta. Setiap tahunnya orang tua siswa mencari sekolah baru untuk anak-anak mereka berharap mendapatkan sekolah yang berkualitas dan unggul. Banyak sekolah menawarkan berbagai fasilitas pendukung, guru yang memadai, dan berbagai tawaran lain agar orang tua siswa menyekolahkan anak-anaknya ke sekolah mereka [7].

Kabupaten Banyumas merupakan kabupaten yang terdapat di Provinsi Jawa Tengah. Di kabupaten ini terdapat 27 kecamatan dan 117 sekolah menengah. Di sekolah menengah yang ada di kabupaten ini terdapat jumlah peserta didik sebanyak 37.917 siswa, 3.383 guru, 830 pegawai, 1.989 ruang kelas, 343 ruang laboratorium, dan 99 perpustakaan. Dari 117 sekolah tersebut tersebar jumlah peserta didik, guru, dan

pegawai yang tidak merata. Kualitas dari tiap sekolah tersebut juga beragam dan tiap sekolah ingin meningkatkan kualitas dari potensi mereka. Dana yang ada di kabupaten Banyumas terbatas, sehingga perlu suatu cara untuk mengelompokkan terlebih dahulu sekolah-sekolah tersebut menjadi beberapa kelompok menggunakan suatu metode pengelompokan atau clustering. Algoritma X-Means merupakan algoritma yang menyempurnakan kekurangan dari algoritma K-means, dimana algoritma K-Means memiliki kelemahan dalam komputasi yang relative lama dan nilai kluster K yang harus ditentukan oleh user. Rumusan masalah dalam penelitian ini adalah bagaimana membuat kelompok atau kluster menggunakan metode pengelompokan X-Means, sedangkan tujuan dari penelitian ini adalah menghasilkan kelompok atau kluster sebagai rekomendasi pada bagian variable mana yang perlu ditingkatkan agar menjadi sekolah unggul menggunakan metode pengelompokan X-Means.

## II. LITERATURE REVIEW

### A. Analisis Cluster

Analisis Cluster termasuk dalam teknik analisis *multivariate method interdependen* yang mewakili seluruh hubungan interdependensi, tidak ada perbedaan variabel bebas dan variabel tak bebas (*independent and dependent variables*) dalam analisis ini. Teknik ini bertujuan untuk mengatur informasi/ meringkas data dengan cara mengelompokkan objek-objek berdasarkan kesamaan karakteristik tertentu. Dengan demikian, analisis cluster atau *clustering* memiliki tujuan untuk mengelompokkan data dari serangkaian pengamatan ke *subset* (disebut cluster).

Tujuan utama analisis cluster adalah menentukan struktur data dengan cara meletakkan observasi yang mirip dalam satu kelompok. Pengelompokan hasil observasi yang mirip ke dalam satu kelompok didasarkan pada korelasi antar objek atau dapat juga dengan mengukur *proximity* pada ruang dua dimensi sehingga jarak antara dua observasi menunjukkan kesamaan. Langkah berikutnya adalah menentukan bagaimana membentuk cluster dan berapa jumlah cluster yang akan dibentuk [2]. Atau dalam istilah sederhananya yaitu mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya.

Analisis cluster mengklasifikasi objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain yang berbeda dalam cluster yang sama. Kelompok-kelompok yang terbentuk memiliki homogenitas internal yang tinggi dan heterogenitas eksternal yang tinggi. Berbeda dengan teknik multivariat lainnya, analisis ini tidak mengestimasi set variabel secara empiris. Sebaliknya menggunakan set set variabel yang ditentukan oleh peneliti iut sendiri. Fokus dari analisis cluster adalah membandingkan objek berdasarkan set variabel, hal inilah yang menyebabkan para ahli mendefinisikan set variabel sebagai tahap kritis dalam analisis cluster. Set variabel cluster adalah satu set variabel yang mempresentasikan karakteristik yang dipakai objek-objek.

Solusi analisis cluster bersifat tidak unik, anggota cluster untuk tiap penyelesaian atau solusi tergantung pada beberapa elemen prosedur dan beberapa solusi yang berbeda dapat diperoleh dengan mengubah satu elemen atau lebih. Solusi cluster secara keseluruhan bergantung pada variabel-variabel yang digunakan sebagai dasar untuk menilai kesamaan. Penambahan atau pengurangan variabel-variabel yang relevan dapat mempengaruhi substansi hasil analisis cluster. Analisis cluster merupakan salah satu analisis yang berguna sebagai peringkas data. Dalam meringkas data dapat dilakukan dengan jalan mengelompokkan objek-objek berdasarkan kesamaan karakteristik tertentu di antara objek-objek yang hendak diteliti.

Analisis cluster mengelompokkan sejumlah  $n$  objek berdasarkan  $p$  varia yang secara relatif mempunyai kesamaan karakteristik di antara objek-objek tersebut, sehingga keragaman di dalam satu kelompok tersebut lebih kecil dibandingkan keragaman antar kelompok tersebut lebih kecil dibandingkan keragaman antar kelompok. Objek dapat berupa barang, jasa, tumbuhan, binatang dan orang (responden, konsumen, atau yang lainnya). Objek tersebut akan diklasifikasi ke dalam satu atau lebih cluster (kelompok) sehingga objek-objek yang berada dalam satu cluster (kelompok) akan mempunyai kemiripan atau kesamaan [9].

Dalam analisis penelitian pendidikan, data untuk proses *clustering* dapat berupa data siswa, orang tua, jenis kelamin atau nilai ujian. *Clustering* merupakan metode penting untuk memahami dan utilitas dari cluster

dalam penelitian pendidikan. Analisis cluster dalam penelitian pendidikan dapat digunakan untuk data eksplorasi, cluster konfirmasi dan pengujian hipotesis. Data eksplorasi digunakan ketika ada sedikit informasi tentang sekolah atau siswa yang akan dikelompokkan bersama-sama. Ini bertujuan untuk menemukan setiap cluster yang berarti unit berdasarkan langkah-langkah pada satu set variabel respons. Konfirmasi cluster digunakan untuk mengonfirmasikan hasil cluster yang dilaporkan sebelumnya pengujian Hipotesis yang digunakan untuk mengatur struktur cluster.

Pengelompokan dilakukan berdasarkan kemiripan (*Similarity*) antar objek. Kemiripan diperoleh dengan meminimalkan jarak antar objek dalam kelompok (*Within-Cluster*) dan memaksimalkan jarak antar kelompok (*Between-Cluster*).

### B. X-Means

Algoritma X-Means merupakan algoritma yang menyempurnakan kekurangan dari algoritma KMeans, dimana algoritma K-means memiliki kelemahan dalam komputasi yang relative lama dan nilai kluster K yang harus ditentukan oleh user. Algoritma pembelajaran ini termasuk dalam kategori *unsupervised learning* dimana komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. Algoritma X-Means mengoptimalkan nilai Bayesian Information Criterion (BIC). Komputasi pada pengoptimalan nilai BIC jauh lebih efisien daripada satu kali iterasi pada algoritma K-Means (Pelleg and Moore, 2000).

### C. Model Searching

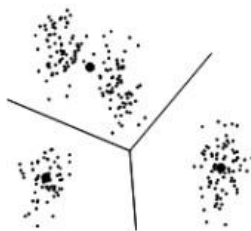
Jika K-Means membutuhkan input dari user untuk nilai k, maka dalam algoritma x-means yang diinputkan oleh user adalah range cluster. Algoritma xmeans dimulai dengan nilai k sama dengan batas bawah (lower bound) dan terus bertambah hingga mencapai batas atas (upper bound) dari range cluster yang sudah diberikan. Selama proses ini berlangsung, jumlah k yang memperoleh skor terbaik dicatat untuk kemudian dikeluarkan ketika proses eksekusi berakhir [5].

Algoritma X-Means terdiri dari dua operasi yang berulang hingga eksekusi berakhir, yaitu:

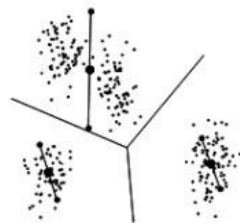
1. Improve – Params
2. Improve – Structure
3. If  $k > k_{max}$ , maka proses eksekusi berhenti dan mengeluarkan model k terbaik yang ditemukan selama proses berlangsung

Improve-Params terdiri dari eksekusi algoritma k-means konvensional hingga mencapai konvergen (tidak terjadi perpindahan point antar cluster).

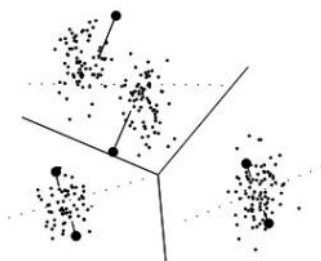
Improve-Structure menemukan jika dan di mana centroid (cluster center) baru harus muncul. Hal ini dicapai dengan membiarkan beberapa centroid dibagi (split) menjadi dua.



Gambar 1. Hasil Klaster Dengan K=3

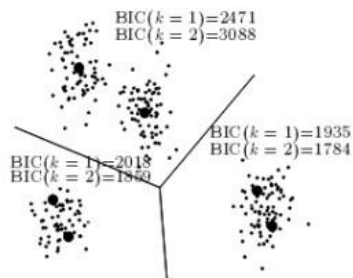


Gambar 2. Pembagian Setiap Centroid Ke Dalam 2 Children



Gambar 3. Langkah Pertama Dalam Local 2 Mean (Pembagian 2 Childeren Baru Di Setiap Centroid)

Asumsi jumlah kluster minimal = 3, maka Gambar 1 menunjukkan hasil dari bagian Improve-Params yaitu eksekusi k-means dengan  $k = 3$ . Langkah berikutnya yaitu Improve-Structure (Gambar 2) dimulai dengan membagi setiap centroid menjadi dua centroid baru (children). Dua hasil pembagian ini bergerak berlawanan arah dalam ukuran region searah dengan vector yang dipilih secara random. Selanjutnya, dalam setiap region centroid awal (parent) dijalankan algoritma k-means secara lokal (dengan  $k = 2$ ) untuk setiap pasang children. Setiap pasang children dibandingkan skor dalam region parentnya. Gambar 3. menunjukkan langkah pertama dari ketiga eksekusi lokal 2-means. Gambar 4. menunjukkan di mana children akhirnya akan berada setelah semua eksekusi lokal 2-means dan dibandingkan dengan nilai sebelumnya [5].



Gambar 4. Hasil Setelah Eksekusi Dari Local 2 Means



Gambar 5. Hasil Akhir Centroid Setelah Pengujian Lokal

Pada tahap ini sebuah uji seleksi model dilakukan pada semua pasang children. Dalam setiap bagian, jika parent dinyatakan memiliki skor yang lebih baik, maka hasil children akan dibuang begitupun sebaliknya (Pelleg dan Dan, 2004). Gambar 5 menunjukkan apa yang terjadi setelah pengujian dilakukan pada tiga pasang children di Gambar 4. Algoritma kemudian meneruskan proses Improve-Params dan Improve-Structure hingga nilai  $k$  sama dengan batas atas range cluster[5].

#### D. BIC Scoring

Pengujian BIC atau BIC scoring dilakukan untuk mengetahui model terbaik setelah eksekusi lokal 2-means berakhir dan untuk mengetahui jumlah cluster terbaik yang dibentuk. Pada dasarnya BIC scoring menggunakan probabilitas posterior untuk memberikan nilai dari model. Formula untuk mengukur posterior sampai normalisasi mengacu pada formula dari Kass dan Wasserman [12] berikut:

$$BIC(M_j) = \hat{l}_j(D) - \frac{P_j}{2} \cdot \log R \quad (1)$$

Dimana  $\hat{l}_j(D)$  adalah fungsi log-likelihood data berdasarkan modek ke- $j$  dan  $P_j$  merupakan jumlah parameter pada  $M_j$ . Dimana diketahui juga sebagai Schwarz Criterion. Kemudian formula untuk menghitung estimasi probabilitas maksimum atau maximum likelihood estimate (MLE) dengan asumsi gaussian adalah:

$$\hat{\sigma}^2 = \frac{1}{R - K} \sum (x_i - \mu_{(i)})^2 \quad (2)$$

Dengan poin probabilitas:

$$\hat{P}(x_i) = \frac{R_{(i)}}{R} \cdot \frac{1}{\sqrt{2\pi\hat{\sigma}^M}} \exp\left(-\frac{1}{2\hat{\sigma}^M} \|x_i - \mu_{(i)}\|^2\right) \quad (3)$$

Likelihood data sebagai berikut:

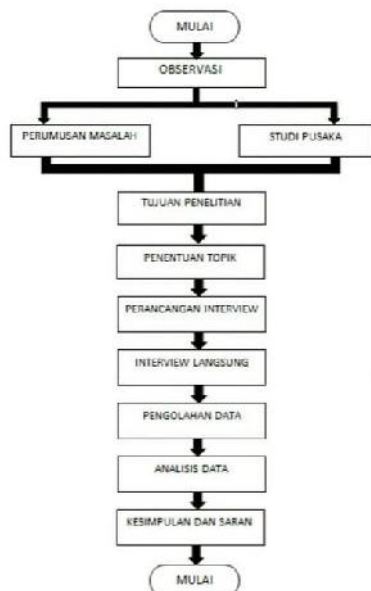
$$l(D) = \log \prod_i P(x_i) = \sum_i \left( \log \frac{1}{2\pi\hat{\sigma}^M} - \frac{1}{2\hat{\sigma}^2} \|x_i - \mu_{(i)}\|^2 + \log \frac{R_{(i)}}{R} \right) \quad (4)$$

Perhitungan hanya berfokus pada set  $D_n$  dari poin untuk setiap centroid  $n$  dan melakukan perhitungan maksimum likelihood sebagai berikut:

$$\hat{l}(D_n) = -\frac{R_n}{2} \log(2\pi) - \frac{R_n \cdot M}{2} \log(\hat{\sigma}^2) - \frac{R_n - K}{2} + R_n \log R_n - R_n \log R \quad (5)$$

Penggunaan formula BIC secara global digunakan ketika X-means memiliki model terbaik. Dan digunakan pada pembagian cluster local 2 means.

### III. RESEARCH METHOD



#### A. Observasi

Dalam langkah observasi ini, peneliti melihat keadaan lapangan secara langsung. Dan dengan melihat keadaan lapangan secara langsung, peneliti dapat menjabarkan langkah-langkah apa yang selanjutnya yang akan dilakukan.

#### B. Perumusan Masalah

Dari beberapa permasalahan yang teridentifikasi selama observasi, penelitian ini menarik satu permasalahan yang ada dan menarik rumusan masalah untuk diteliti dan dilakukan analisis terhadap pengolahan data yang ada. Adapun rumusan permasalahan yang dilakukan pada penelitian ini berlangsung adalah upaya perbaikan kompetensi karyawan yang diharapkan akan dapat meningkatkan kemampuan perusahaan untuk mencapai tujuannya yaitu peningkatan profitabilitas dan sustainabilitas.

#### C. Studi Pustaka

Studi pustaka ialah proses penelusuran / usaha yang dilakukan oleh peneliti untuk menghimpun informasi yang relevan dengan topik atau masalah yang akan atau sedang diteliti. Studi kepustakaan merupakan langkah yang sangat penting dalam metode ilmiah untuk mencari sumber data yang akan mendukung penelitian dan untuk mengetahui sejauh apa pemahaman terhadap ilmu yang berhubungan dengan penelitian telah berkembang, dan sejauh apa terdapat kesimpulan dan degeneralisasi yang pernah dibuat. Dengan adanya studi pustaka, diharapkan pihak lain dapat memahami keseluruhan isi dari laporan tugas akhir ini, terkait pemahaman mengenai topik pembahasan dan tujuan dari penulisan laporan tugas akhir ini.

#### D. Tujuan Penelitian

Tujuan penelitian merupakan rumusan kalimat yang menunjukkan adanya hasil, sesuatu yang diperoleh setelah penelitian selesai, sesuatu yang akan dicapai atau dituju dalam sebuah penelitian.

#### E. Penentuan Topik

Persoalan yang hendak dibuktikan. Dalam hal ini peneliti menentukan satu topik sebagai fokus kegiatan penelitian. Hal itu dilakukan karena permasalahan yang ada biasanya sangat kompleks, sehingga tidak mungkin hanya diteliti dari satu sudut disiplin ilmu saja dan tidak mungkin diteliti dari semua segi secara serentak.

#### *F. Perancangan Interview*

Perancangan interview di buat untuk menyusun apa saja yang akan di tanyakan kepada narasumber untuk memperoleh data yang di inginkan.

#### *G. Interview Langsung*

Interview adalah teknik pengumpulan informasi yang memungkinkan penelitian mempelajari sikap-sikap keyakinan, perilaku, karakteristik seseorang yang bisa terpengaruh oleh sistem yang sudah ada. Dengan menggunakan interview, penelitian ini berupaa mengukur apa yang ditemukan dalam wawancara / interview, selain itu juga untuk menentukan seberapa luas atau terbatasnya sentiment yang diekspresikan dalam suatu wawancara. Penggunaan interview tepat bila:

- a. Responden (orang yang merespons atau menjawab pertanyaan) saling berjauhan antara responden yang satu dengan responden yang lainnya.
- b. Melibatkan sejumlah orang dalam wawancara, sehingga penelitian bisa mengetahui beberapa proporsi suatu kelompok tertentu yang menyetujui atau tidak menyetujui suatu sistematis dari sistem yang sudah ada.
- c. Melakukan studi untuk mengetahui sesuatu dan ingin mencari seluruh pendapat sebelum proyek sistem diberi petunjuk – petunjuk tertentu.
- d. Ingin meyakinkan bahwa masalah – masalah dalam sistem yang ada bisa diidentifikasi dan dibicarakan dalam wawancara tindak lanjut.

Perbedaan pertanyaan dalam wawancara dengan pertanyaan dalam interview memungkinkan adanya interaksi antara pertanyaan dan artinya. Dalam wawancara analisis memiliki peluang untuk menyaring suatu pertanyaan, menetapkan istilah – istilah yang belum jelas, mengubah arus pertanyaan, member respons terhadap pandangan yang rumit dan umumnya bisa mengontrol agar sesuai dengan konteksnya.

Beberapa diantara peluang-peluang diatas juga dimungkinkan dalam interview. Oleh karena itu penelitian ini memberikan pertanyaan-pertanyaan jelas, masuk akal, dan pertanyaan – pertanyaan dari responden diantisipasi agar susunan pertanyaan direncanakan secara mendetail. Jenis – jenis pertanyaan dalam interview adalah:

- a. Pertanyaan terbuka; pertanyaan – pertanyaan yang member pilihan – pilihan respons terbuka kepada responden. Pada pertanyaan terbuka antisipasilah jenis respons yang akan muncul. Respons yang diterima harus tetap bisa diterjemahkan dengan benar.
- b. Pertanyaan tertutup; pertanyaan – pertanyaan yang membatasi atau menutup pilihan – pilihan respons yang tersedia bagi responden.

#### *8. Pengolahan Data*

Pada pengolahan data penelitian ini diarahkan untuk mengolah informasi dan data yang diperoleh dari berbagai sumber yang ada. Data yang diperoleh ditindaklanjuti melalui serangkaian metode dan juga pengujian yang disesuaikan dengan kebutuhan dan juga keperluan.

#### *9. Analisis Data*

Pada tahapan ini akan dilakukan analisis dan pembahasan dari hasil pengolahan data yang telah dilakukan pada tahapan sebelumnya. Pengolahan data yang ada akan dikur kompetensi pada saat ini, kemudian akan

dilakukan pengukuran kompetensi kesenjangan dengan kompetensi yang diharapkan, dan usulan tindakan untuk memperbaiki kesenjangan kompetensi tersebut.

#### IV. RESULTS AND DISCUSSION

Hasil dari penggunaan X-means dalam pengelompokan cluster 188 sekolah di kabupaten banyumas menghasilkan empat cluster dengan jarak rata-rata *Euclidean distance* cluster adalah 1.875. dengan nilai Davis Bouldin Index (DBI) sebesar 0.846. Dengan nilai DBI yang masih kurang dari satu, maka hasil 4 cluster masih terbilang cukup ideal untuk digunakan dalam menentukan pengelompokan sekolah di banyumas berdasarkan tujuh kriteria. Ketujuh kriteria tersebut yaitu jumlah peserta didik, jumlah rombongan belajar, jumlah guru, jumlah pegawai, jumlah ruang kelas, jumlah perpustakaan dan jumlah laboratorium. Pada cluster pertama terdapat 26 sekolah dengan keterangan 20 SMK Swasta, 1 SMK Negeri, 3 SMA Negeri dan 2 SMA Swasta. 28 sekolah pada cluster kedua dengan keterangan 11 SMA Negeri, 7 SMK Negeri dan 10 SMK Swasta. 21 sekolah pada cluster ketiga dengan keterangan 19 SMK Swasta dan 2 SMA Swasta. Cluster terakhir beranggotakan 43 sekolah dengan keterangan 19 SMA Swasta, 1 SMK Negeri dan 23 SMK Swasta. Dengan masing-masing kriteria untuk setiap cluster ditunjukkan pada Tabel 1.

TABEL I  
KRITERIA RATA-RATA DATA UNTUK SETIAP CLUSTER

Cluster	Peserta Didik	Rombel	Guru	Pegawai	R. Kelas	R. Lab	R. Perpus
C1	687	24	34	7	22	3	1
C2	1321	38	59	16	33	5	1
C3	153	7	10	2	7	1	0
C4	159	8	14	3	8	2	1

Dari data di Tabel 1, maka pengelompokan data sekolah dapat dikelompokkan menjadi empat kelompok dengan kelompok pertama yaitu kelompok unggul pada cluster 2 (C2), kelompok baik pada cluster 1 (C1), kelompok cukup pada cluster 4 (C4) dan kelompok kurang pada cluster 3 (C3). Anggota sekolah yang ada pada C2 memiliki nilai rata-rata kriteria yang paling tinggi sehingga dikategorikan sebagai sekolah 'unggul'. Sedangkan anggota sekolah pada C1 memiliki nilai rata-rata kriteria dibawah C2 akan tetapi masih diatas C3 dan C4 sehingga dikategorikan sebagai sekolah 'baik'. C3 Karena memiliki nilai rata-rata kriteria yang paling rendah dibandingkan cluster lainnya maka dikategorikan sebagai sekolah 'kurang'. C4 memiliki nilai rata-rata kriteria di atas C3, sehingga dikategorikan pada sekolah 'cukup'. Akan tetapi karena C4 memiliki anggota cluster yang paling banyak hampir dua kali dari jumlah anggota cluster lainnya.

#### V. CONCLUSION

Penggunaan X-Means dalam pengelompokan SMA dan SMK di Banyumas memberikan hasil yang cukup baik pada empat cluster dengan nilai DBI sebesar 0.846. Empat cluster tersebut adalah cluster unggul dengan 28 anggota, cluster baik dengan 26 anggota, cluster cukup dengan 43 anggota dan cluster kurang sebanyak 21 anggota. Dengan hasil tersebut maka dapat diambil kesimpulan bahwa tingkat pendidikan menengah Kabupaten Banyumas masih masuk ke dalam cluster cukup.

#### ACKNOWLEDGMENT

Peneliti mengucapkan terimakasih atas pendanaan penelitian oleh ristekdikti dengan nomor xxxx tahun 2019

Comment [ABA1]: Belum lengkap



#### REFERENCES

- [1] Everitt. 1993. Cluster Analysis. Edward Arnold. London
- [2] Ghozali. Imam, 2009. Aplikasi Analisis Multivariate dengan Program SPSS. Badan Penerbit Universitas Diponegoro. Vol. 100-125. Semarang.
- [3] Mumiati S, 2012. Pengertian Fungsi dan Jenis Lingkungan Pendidikan. Lingpendidikan.com. Palembang.
- [4] Pengkab Banyumas, 2017. Perubahan Rencana Strategis Tahun 2013-2018 Dinas Sumber Daya Air dan Bina Marga Kabupaten Banyumas menjadi Rencana Strategis Tahun 2017/2018 Dinas Pekerjaan Umum Kabupaten Banyumas. Dinas Pekerjaan Umum Pemerintah Kabupaten Banyumas. Banyumas.
- [5] Pelleg, Dan dan Moore, Andrew. 2000. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. Pittsburgh: Carnegie Mellon University, 2000.
- [6] Pelleg, Dan. 2004. K-means and KD-trees resources. Dan Pelleg's Home Page. [Online] National Science Foundation, Agustus 13, 2004. [Cited: Januari 3, 2009.] <http://www.cs.cmu.edu/~dpelleg/K-means.html>.
- [7] Purnama D. 2010. Cermat Memilih Sekolah Menengah yang Tepat. Gagas Media. Jakarta.
- [8] Russel, S. and Norvig, P. 2010. Artificial Intelligence: A Modern Approach. 3rd Edition, Prentice Hall, Upper Saddle River.
- [9] Soraya, Yani, 2011, Perbandingan Kinerja Metode Single Linkage, Metode Complete Linkage, dan Metode K-Means dalam Analisis Cluster, Jurusan Matematika Fakultas MIPA Universitas Negeri Semarang, Semarang.
- [10] Supranto, J. 2004, Analisis Multivariate: Arti dan Interpretasi, Rineka Cipta, Jakarta.
- [11] Tan P, Steinbach M, Kumar V. 2006. Introduction to Data Mining. Pearson AddisonWeasley
- [12] Robert E. Kass and Larry Wasserman. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, 90(431):928-934, 1995.