



Perbandingan Hasil Analisis Klusterisasi Data Bimbingan Konseling Peserta Didik Tingkat Menengah Pertama sebagai Identifikasi Kinerja Akademik

Ratna Nur Tiara Shanty¹, Rizki Nurhidayat², Dwi Cahyono³

^{1,2,3}*Teknik Informatika, Universitas Dr.Soetomo, Surabaya, Indonesia*

STATUS ARTIKEL

Dikirim 03 September 2020
Direvisi 30 September 2020
Diterima 3 Oktober 2020

Kata Kunci:

*Klusterisasi, K-Means, Hierarchical,
Orange Visual Programming,
Bimbingan Konseling, Kinerja
Akademik*

ABSTRAK

Bimbingan Konseling (BK) merupakan sarana penting pada sekolah yang berperan dalam identifikasi dan pengembangan kepribadian, kehidupan sosial, kegiatan belajar, serta potensi karir bagi peserta didik. Pada SMP Negeri 12 Surabaya, rasio jumlah guru BK dengan siswanya adalah sekitar 1 banding 458. Maka, dibutuhkan efisiensi dalam penanganan dengan membagi siswa menjadi beberapa kelompok. Hubungan antar atribut penilaian evaluasi diri, riwayat pelanggaran, dengan nilai hasil belajar yang diperoleh seorang siswa tidak selalu berbanding lurus. Pada penelitian ini dilakukan perbandingan hasil analisis pola kinerja siswa menggunakan metode K-Means dan Hierarchical Single Linkage, Average Linkage dan Complete Linkage. Pola kinerja siswa dianalisis berdasarkan atribut-atribut evaluasi diri pada kriteria permasalahan pribadi, sosial, belajar, dan karir, serta hasil prestasi akademik, dan riwayat pelanggaran yang pernah dilakukan oleh setiap siswa. Uji coba dilakukan dengan perangkat lunak Orange Visual Programming. Hasil pada penelitian ini disimpulkan bahwa metode klusterisasi menghasilkan kelompok-kelompok siswa dengan kecenderungan nilai fitur yang berbeda, dan metode Hierarchical Single Linkage memiliki performa klusterisasi terbaik dengan rasio sebesar 0.004435.

1. PENDAHULUAN

Bimbingan Konseling (BK) merupakan sarana penting pada sekolah yang berperan dalam identifikasi dan pengembangan kepribadian, kehidupan sosial, kegiatan belajar, serta potensi karir bagi peserta didik (Triyanto, 2017). Saat ini, lembaga BK yang ada di lingkungan Sekolah Menengah Pertama (SMP) Negeri 12 Surabaya melaksanakan peranannya dengan cara menggali permasalahan pada kehidupan pribadi, identifikasi potensi karir, pemantauan prestasi akademik, hingga riwayat pelanggaran yang pernah dilakukan oleh setiap siswa. Namun, keseluruhan proses tersebut dikerjakan dengan cara manual, yaitu Guru BK sebagai konselor, mencatat hasil riwayat bimbingan setiap siswa ke dalam sebuah buku besar. Hal tersebut merupakan sebuah tantangan yang besar untuk suatu sekolah dalam mengidentifikasi kinerja akademik siswanya secara menyeluruh.

Teknologi informasi berperan penting dalam segala sektor kehidupan termasuk di bidang pendidikan. Tidak sedikit penelitian yang telah mengimplementasikan ilmu teknologi informasi yang menunjang proses bimbingan konseling oleh Guru BK. Penelitian-penelitian tersebut berkontribusi melalui perancangan sistem informasi (Juanita & Wibisono, 2019), (Hanafi, Wibowo, & Putra, 2020), juga pada perancangan basis data (Handayani & Putri, 2017). Namun,

diperlukan ekstraksi pengetahuan yang tersembunyi dari tumpukan data yang dihasilkan dari data hasil transaksi suatu sistem informasi. Ekstraksi pengetahuan dapat diselesaikan menggunakan teknik data mining yang merupakan bagian dari ilmu teknologi informasi.

Pada SMP Negeri 12 Surabaya, rasio jumlah guru BK dengan siswanya adalah sekitar 1 banding 458. Maka, dibutuhkan efisiensi dalam penanganan dengan membagi siswa menjadi beberapa kelompok. Cara yang paling sederhana adalah dengan menggunakan ambang batas nilai-nilai yang telah ditentukan pada setiap atribut. Namun, hubungan antar atribut penilaian evaluasi diri, riwayat pelanggaran, dengan nilai hasil belajar yang diperoleh seorang siswa tidak selalu berbanding lurus. Artinya, pada seorang siswa bukan berarti jika hasil evaluasi dirinya baik, maka tidak pernah melanggar dan hasil nilai belajarnya selalu tinggi, begitu juga sebaliknya. Klasterisasi sebagai salah satu area pada data mining dapat menjadi sebuah solusi dari permasalahan tersebut. Klasterisasi bermanfaat untuk menghasilkan kelompok-kelompok dari sekumpulan data berdasarkan kemiripan atribut yang dimiliki oleh siswa. Beberapa penelitian menggunakan metode K-Means (Borgavakar & Shrivastava, 2017) dan metode Hierarchical (Rana & Garg, 2016) untuk mengidentifikasi kinerja akademik. Masing-masing metode tersebut tentunya memiliki karakteristik yang berbeda.

Pada penelitian ini dilakukan perbandingan hasil analisis pola kinerja siswa menggunakan teknik klasterisasi. Pola kinerja siswa dianalisis berdasarkan atribut-atribut evaluasi diri pada kriteria permasalahan pribadi, sosial, belajar, dan karir. Kemudian atribut hasil prestasi akademik, dan riwayat pelanggaran yang pernah dilakukan oleh setiap siswa juga akan menjadi titik fokus pada penelitian ini. Metode klasterisasi yang dibandingkan adalah K-Means, dan metode Hierarchical dengan perhitungan single linkage, average linkage, dan complete linkage. Uji coba dilakukan dengan perangkat lunak Orange Visual Programming untuk memperlihatkan perbedaan hasilnya. Hasil dari penelitian ini diharapkan dapat bermanfaat untuk mengetahui metode klasterisasi yang memiliki performa yang baik, dan nantinya dapat dikembangkan menjadi sebuah aplikasi sistem yang dapat digunakan oleh pihak lembaga BK pada sekolah menengah pertama, khususnya pada SMP Negeri 12 Surabaya.

2. METODE

2.1. *Kebutuhan Data*

Data yang diproses pada penelitian ini adalah data yang berasal dari kegiatan bimbingan konseling di lingkungan SMP Negeri 12 Surabaya. Data memiliki 3 bagian utama yang digambarkan pada Gambar 2.1, yaitu data evaluasi diri, data riwayat pelanggaran, dan data nilai rapor sebagai hasil penilaian akademik. Masing-masing siswa tentunya akan memiliki nilai yang berbeda pada setiap data, bergantung pada hasil evaluasi diri yang diisi, pelanggaran apa saja yang dilakukan, dan bagaimana pencapaian hasil belajar pada periode tertentu.

A. Data Evaluasi Diri

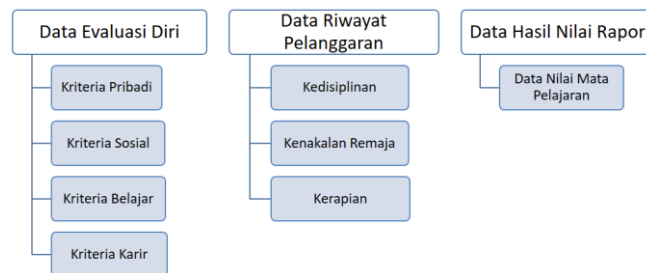
Data pertama yaitu data evaluasi diri yang berasal dari proses penggalian informasi peserta didik yang akan digunakan untuk mengidentifikasi permasalahan pada kriteria pribadi, sosial, belajar dan karir. Data ini didapatkan dari hasil pengisian kuesioner yang dijawab oleh siswa. Potongan uraian pertanyaan pada kuesioner data evaluasi diri dapat dilihat pada Tabel 2.1.

B. Data Riwayat Pelanggaran

Data kedua adalah data pelanggaran yang didapatkan dari proses pencatatan oleh BK. Poin-poin pelanggaran mempunyai nilai sesuai dengan jenis dan tingkat pelanggarannya. Potongan jenis-jenis pelanggaran diuraikan pada Tabel 2.2.

C. Data Rapor Hasil Penilaian Belajar

Data ketiga adalah data hasil penilaian rapor pada setiap jenjang kelas yang akan dijadikan sebagai identifikasi awal prediksi performa akademik siswa selanjutnya.



Gambar 2.1 Bagan Kebutuhan Data

Tabel 2.1 Potongan Daftar Pernyataan pada Kuesioner Evaluasi Diri

Kriteria	Pernyataan	Nilai
Pribadi	Saya belum bersungguh-sungguh beribadah pada Tuhan YME	1
Pribadi	Kadang-kadang perbuatan saya tidak sesuai dengan yang diucapkan	1
Pribadi	Saya merasa kurang memiliki rasa tanggung jawab	1
Pribadi	Saya gampang marah tanpa tahu penyebabnya	1
Pribadi	Saya merasa kurang mendapatkan perhatian dari orang tua	1
Sosial	Saya belum tahu tentang bullying dan cara mengikapinya	2
Sosial	Saya merasa malu jika bergaul dengan teman yang beda jenis kelamin	2
Sosial	Orang tua saya tidak peduli dengan kegiatan belajar saya	2
Sosial	Saya merasa tidak nyaman kalau belajar di rumah sendiri	2
Belajar	Saya belum terbiasa belajar kelompok. biasanya saya selalu belajar sendiri	3
Belajar	Saya belum paham cara yang baik belajar di sekolah baru (SMP/MTs)	3
Belajar	Saya merasa pesimis bisa naik kelas	3
Karir	Saya belum mempunyai cita-cita yang pasti	4
Karir	Saya belum banyak tahu tentang jenis-jenis pekerjaan di masyarakat	4
Karir	saya merasa belum paham hubungan antara hobi. bakat. minat dan kemampuan	4

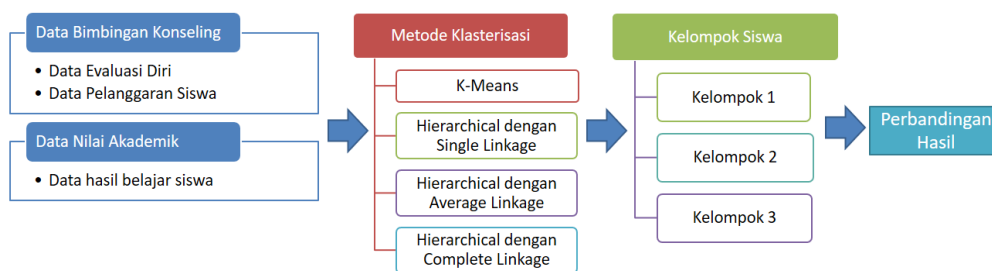
Tabel 2.2 Potongan Daftar Jenis Pelanggaran

Nama Pelanggaran	Poin
Melakukan kecurangan saat ulangan	5
Tidak melaksanakan tugas yang di beri sekolah/tidur pada saat pelajaran berlangsung/tidak meperhatikan guru	10
Terlambat mengikuti ulangan dan ujian	20
Membawa dan melihat gambar porno atau sejenisnya yang merendahkan martabat	30
Tidak sopan kepada guru/karyawan dalam komunikasi (omong kotor,memaki,mengumpat dll)	30
Berciuman/mencium	35
Melakukan pengerusakan sarana dan prasarana secara perorangan	50
Berkelahi dengan teman atau orang luar	50
Menjadi dalang perkelahian	150

Nama Pelanggaran	Poin
Terlibat perbuatan pidana	200
Melakukan tindak asusila	200
Mencuri	200
Memukul guru dan karyawan	300
Melakukan hubungan sex /hamil	300

2.2. Blok Diagram

Secara umum prosesnya digambarkan pada Gambar 2.2. Data lembaga bimbingan konseling dan data nilai akademik siswa diproses oleh empat perhitungan klasterisasi yaitu K-Means, Hierarchical dengan perhitungan *single linkage*, *average linkage*, dan *complete linkage*. Kelompok-kelompok yang dihasilkan dari proses klasterisasi tersebut akan dievaluasi dan dibandingkan hasil pola pengelompokkannya.



Gambar 2.2 Blok Diagram Proses Keseluruhan

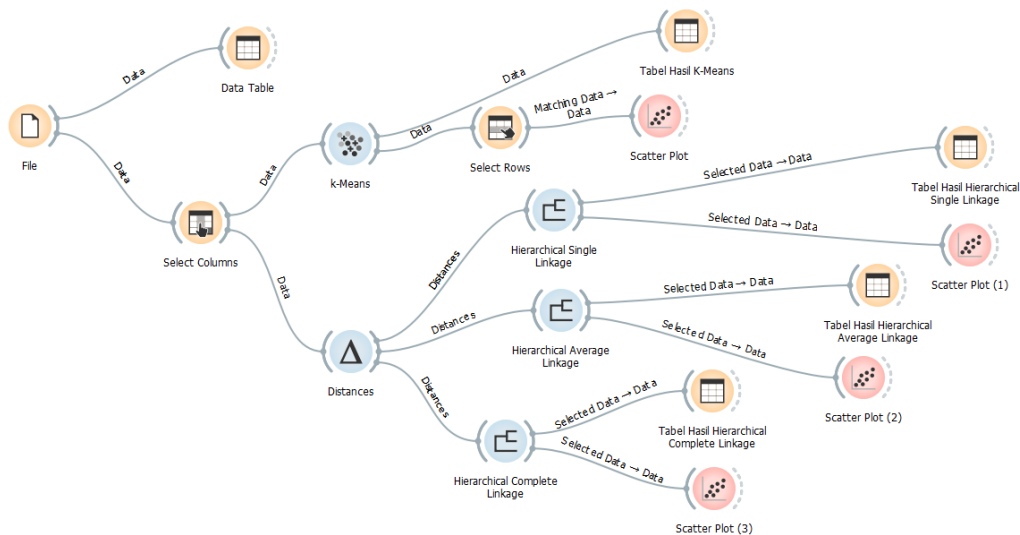
2.3. Perangkat Lunak Penunjang

Orange Visual Programming digunakan sebagai perangkat lunak data mining untuk mencapai tujuan pada penelitian ini. Perangkat lunak ini berbasis bahasa pemrograman Python. Agar dapat melakukan analisis pola data, pada perangkat lunak ini disediakan widget-widget yang harus dirangkai sesuai dengan tahapan pada data mining. Hasil dari rangkaian kemudian disebut juga dengan istilah *Work flow*.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini, *Work flow* dirangkai sesuai dengan yang ditunjukkan pada Gambar 3.1. Serangkaian proses dimulai dari pembacaan sekumpulan data, melakukan pemrosesan awal untuk seleksi atribut mana saja yang diproses sebagai fitur dan yang hanya digunakan sebagai meta data. Kemudian, diproses menggunakan algoritma klasterisasi K-Means, Hierarchical dengan perhitungan *single linkage*, *average linkage*, dan *complete linkage*.

Widget File pada Orange Visual Programming, dapat digunakan untuk melakukan pembacaan sekumpulan data. Tahapan selanjutnya setelah proses pembacaan data adalah proses seleksi fitur untuk kebutuhan klasterisasi. Fitur yang diproses meliputi hasil akumulasi dari nilai evaluasi diri, poin pelanggaran, dan nilai akademik. Sedangkan Nomor Induk Siswa (NIS) dan Nama Siswa hanya digunakan sebagai meta data. Pengaturan jarak metrik pada penelitian ini menggunakan Euclidean dengan menghitung jarak kemiripan antar baris data. Adapun pengaturan jumlah kelompok yang dibentuk pada penelitian ini adalah sebanyak 3 kelompok.



Gambar 3.1 Work Flow pada Aplikasi Orange Visual Programming

3.1. Analisis Klaster

A. Klasterisasi dengan K-Means

Hasil klasterisasi dengan metode K-Means ditunjukkan oleh Gambar 3.2, dengan hasil *scatter plot* yang ditampilkan Gambar 3.6. Pada hasil *scatter plot* tersebut terlihat bahwa K-Means berhasil memisahkan klaster (cluster) dengan baik pada atribut nilai rapor terhadap poin pelanggaran.

B. Klasterisasi dengan Hierarchical Single Linkage

Hasil klasterisasi dengan metode Hierarchical Single Linkage ditunjukkan oleh Gambar 3.3, dengan hasil *scatter plot* yang ditampilkan pada Gambar 3.7. Pada hasil *scatter plot* tersebut terlihat bahwa Hierarchical Single Linkage berhasil memisahkan cluster dengan baik untuk atribut poin pelanggaran terhadap poin evaluasi diri.

C. Klasterisasi dengan Hierarchical Average Linkage

Hasil klasterisasi dengan metode Hierarchical Average Linkage ditunjukkan oleh Gambar 3.4, dengan hasil *scatter plot* yang ditampilkan pada Gambar 3.8. Pada hasil *scatter plot* tersebut terlihat bahwa Hierarchical Average Linkage berhasil memisahkan cluster dengan baik untuk atribut poin pelanggaran terhadap poin evaluasi diri.

D. Klasterisasi dengan Hierarchical Complete Linkage

Hasil klasterisasi dengan metode Hierarchical Complete Linkage ditunjukkan oleh Gambar 3.5, dengan hasil *scatter plot* yang ditampilkan pada Gambar 3.9. Pada hasil *scatter plot* tersebut terlihat bahwa Hierarchical Complete Linkage berhasil memisahkan cluster dengan baik untuk atribut poin pelanggaran terhadap nilai rapor.

	NIS	Nama Siswa	Cluster	Silhouette	Evaluasi Diri	Poin Pelanggaran	Nilai Rapor
5			C1	0.454014	92	0	70
14			C1	0.546862	92	0	64.6364
15			C1	0.599017	88	0	54.8182
17			C1	0.543387	92	50	64.6364
18			C1	0.604922	18	0	64.6364
20			C1	0.615101	14	50	63.7273
25			C1	0.615622	16	50	64
6			C2	0.632774	92	150	84.3636
9			C2	0.643127	86	300	80
24			C2	0.629899	88	150	86
1			C3	0.705761	88	0	86.1818
2			C3	0.695312	88	20	83.2727
3			C3	0.691579	92	0	82.7273
4			C3	0.705479	92	0	86.5455
7			C3	0.706295	92	5	86.3636
8			C3	0.690156	86	37	86.2727
10			C3	0.689439	92	35	87.3636
11			C3	0.701409	92	5	90.7273
12			C3	0.691891	92	10	92.7273
13			C3	0.702277	92	0	90.5455
16			C3	0.588915	28	0	93.0909
19			C3	0.694423	78	25	86.4545
21			C3	0.696049	88	5	83.0909
22			C3	0.632077	51	5	87.0909
23			C3	0.607744	87	0	75.5455

Gambar 3.2 Hasil Klasterisasi dengan Metode K-Means

	NIS	Nama Siswa	Cluster	Evaluasi Diri	Poin Pelanggaran	Nilai Rapor
18			C1	18	0	64.6364
20			C1	14	50	63.7273
25			C1	16	50	64
9			C2	86	300	80
1			C3	88	0	86.1818
2			C3	88	20	83.2727
3			C3	92	0	82.7273
4			C3	92	0	86.5455
5			C3	92	0	70
6			C3	92	150	84.3636
7			C3	92	5	86.3636
8			C3	86	37	86.2727
10			C3	92	35	87.3636
11			C3	92	5	90.7273
12			C3	92	10	92.7273
13			C3	92	0	90.5455
14			C3	92	0	64.6364
15			C3	88	0	54.8182
16			C3	28	0	93.0909
17			C3	92	50	64.6364
19			C3	78	25	86.4545
21			C3	88	5	83.0909
22			C3	51	5	87.0909
23			C3	87	0	75.5455
24			C3	88	150	86

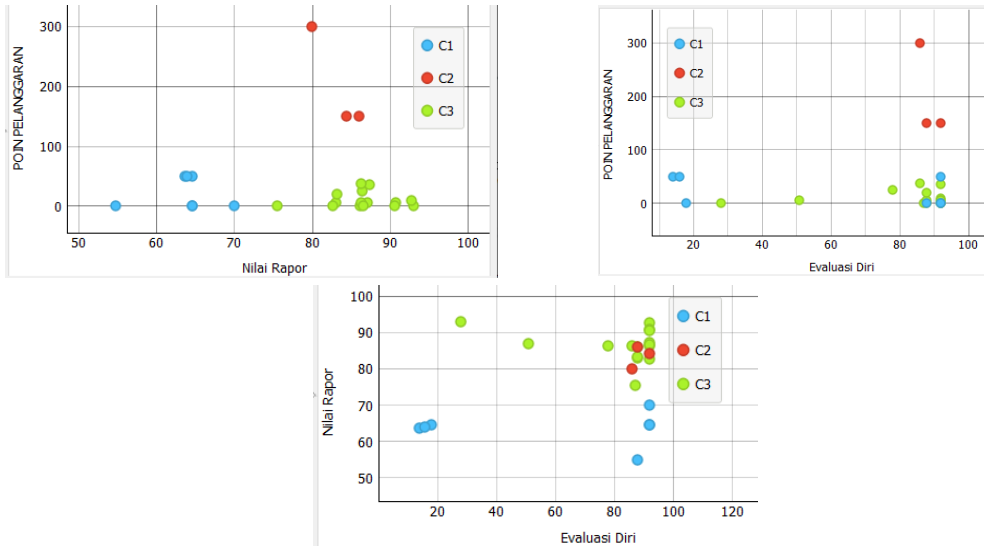
Gambar 3.3 Hasil Klasterisasi dengan Metode Hierarchical Single Linkage

	NIS	Nama Siswa	Cluster	Evaluasi Diri	Poin Pelanggaran	Nilai Rapor
18			C1	18	0	64.6364
20			C1	14	50	63.7273
25			C1	16	50	64
1			C2	88	0	86.1818
2			C2	88	20	83.2727
3			C2	92	0	82.7273
4			C2	92	0	86.5455
5			C2	92	0	70
7			C2	92	5	86.3636
8			C2	86	37	86.2727
10			C2	92	35	87.3636
11			C2	92	5	90.7273
12			C2	92	10	92.7273
13			C2	92	0	90.5455
14			C2	92	0	64.6364
15			C2	88	0	54.8182
16			C2	28	0	93.0909
17			C2	92	50	64.6364
19			C2	78	25	86.4545
21			C2	88	5	83.0909
22			C2	51	5	87.0909
23			C2	87	0	75.5455
6			C3	92	150	84.3636
9			C3	86	300	80
24			C3	88	150	86

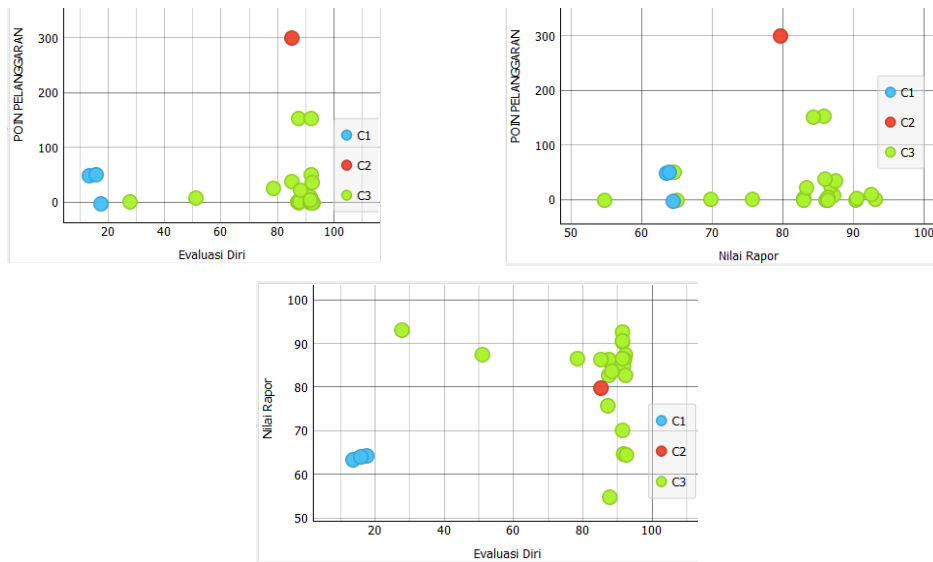
Gambar 3.4 Hasil Klasterisasi dengan Metode Hierarchical Average Linkage

	NIS	Nama Siswa	Cluster	Evaluasi Diri	Poin Pelanggaran	Nilai Rapor
6			C1	92	150	84.3636
9			C1	86	300	80
24			C1	88	150	86
1			C2	88	0	86.1818
2			C2	88	20	83.2727
3			C2	92	0	82.7273
4			C2	92	0	86.5455
7			C2	92	5	86.3636
8			C2	86	37	86.2727
10			C2	92	35	87.3636
11			C2	92	5	90.7273
12			C2	92	10	92.7273
13			C2	92	0	90.5455
16			C2	28	0	93.0909
19			C2	78	25	86.4545
21			C2	88	5	83.0909
22			C2	51	5	87.0909
5			C3	92	0	70
14			C3	92	0	64.6364
15			C3	88	0	54.8182
17			C3	92	50	64.6364
18			C3	18	0	64.6364
20			C3	14	50	63.7273
23			C3	87	0	75.5455
25			C3	16	50	64

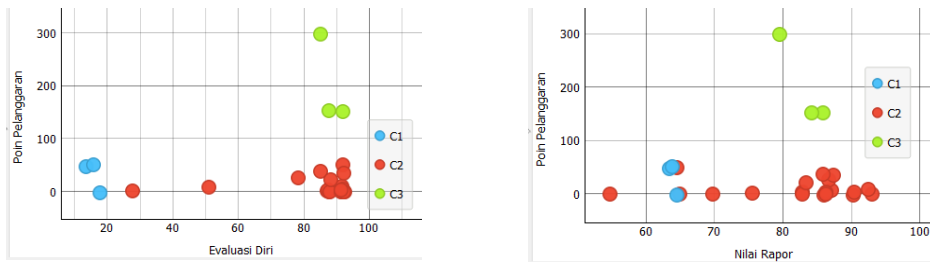
Gambar 3.5 Hasil Klasterisasi dengan Metode Hierarchical Complete Linkage

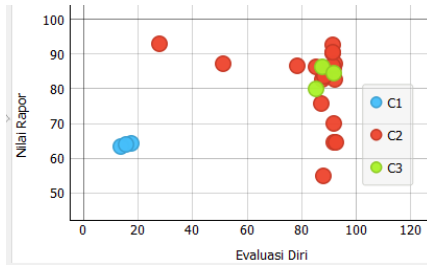


Gambar 3.6 Scatter Plot K-Means

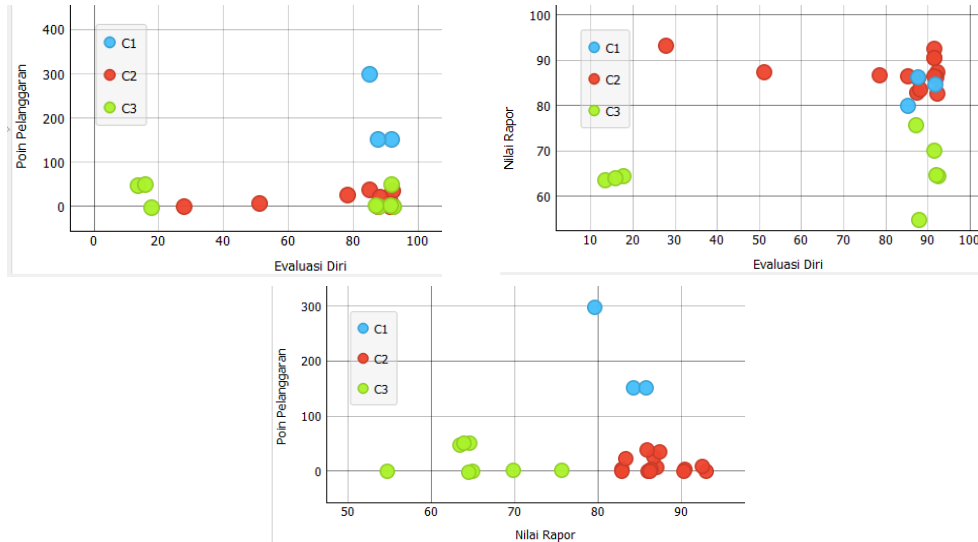


Gambar 3.7 Scatter Plot Hierarchical Single Linkage





Gambar 3.8 Scatter Plot Hierarchical Average Linkage



Gambar 3.9 Scatter Plot Hierarchical Complete Linkage

3.2. Evaluasi menggunakan Varians:

Hasil pengelompokan yang baik adalah yang memiliki varians dalam cluster yang rendah, dan varians antar cluster yang tinggi. Varians adalah tingkat keberagaman pada sekumpulan data. Untuk dapat menghitung varians dalam cluster dan varians antar cluster, maka perlu dihitung terlebih dahulu varians pada cluster tersebut. Adapun rumus untuk menghitung varians pada suatu cluster ditunjukkan pada formula 3.1.

$$v_c^2 = \frac{1}{n_c - 1} \sum_{i=1}^{n_c} (d_i - \bar{d}_i)^2, \quad (3.1)$$

dimana v_c^2 adalah varians pada cluster c ; c adalah cluster, n_c adalah jumlah data pada cluster c , d_i adalah data ke- i pada suatu cluster, dan \bar{d}_i adalah rata-rata data pada suatu cluster. Sehingga rumus untuk menghitung varians dalam cluster dituliskan pada formula 3.2.

$$v_w = \frac{1}{N - k} \sum_{i=1}^k (n_i - 1) \cdot v_i^2, \quad (3.2)$$

dimana v_i^2 adalah varians pada tiap cluster ke- i , n_i adalah jumlah data pada cluster ke- i , k adalah jumlah cluster, dan N adalah jumlah semua data. Selanjutnya, rumus untuk menghitung varians antar cluster adalah ditunjukkan pada formula 3.3.

$$v_b = \frac{1}{k - 1} \sum_{i=1}^k n_i (\bar{d}_i - \bar{d})^2, \quad (3.3)$$

dimana \bar{d} adalah rata-rata semua data pada tiap fitur. Adapun hasil perhitungan varians dalam klaster, antar klaster dan rasionya ditunjukkan pada Tabel 3.1.

Tabel 3.1 Hasil Perhitungan Varians untuk Setiap Metode

Metode	Varians		
	Dalam Kluster (v_w)	Antar Kluster (v_b)	Rasio (v_w/v_b)
K-Means	1658.993	49077.9067	0.03380325
Hierarchical Single Linkage	2198.364	495709.7	0.004435
Hierarchical Average Linkage	1274.172	119261.1	0.010684
Hierarchical Complete Linkage	1702.734	48596.75	0.035038

4. KESIMPULAN

Dari hasil uji coba yang telah dilakukan, maka pada penelitian ini dapat disimpulkan sebagai berikut:

1. Setiap metode klasterisasi menghasilkan kelompok-kelompok siswa dengan kecenderungan nilai fitur yang berbeda.
2. Berdasarkan evaluasi perhitungan tingkat keberagaman data, bahwa hasil klasterisasi yang baik adalah yang memiliki rasio varians dalam kluster dibanding varians antar kluster terkecil. Sehingga dari Tabel 3.1, disimpulkan bahwa metode Hierarchical Single Linkage memiliki performa klasterisasi terbaik dengan rasio sebesar 0.004435.

5. UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada lembaga Bimbingan Konseling SMP Negeri 12 Surabaya yang telah memberi dukungan survei kondisi dan data.

6. DAFTAR PUSTAKA

- Borgavakar, S., & Shrivastava, A. (2017). Evaluating Student's Performance using K-Means Clustering. *International Journal of Engineering Research & Technology (IJERT)*, 114-116.
- Hanafi, R., Wibowo, N., & Putra, A. (2020). SISTEM INFORMASI BIMBINGAN KONSELING BERBASIS WEB (STUDI KASUS : SMK NEGERI 1 MOJOAGUNG). *Jurnal Informatika dan Sistem Informasi (JIFoSI)*, 38-51.
- Handayani, F., & Putri, M. (2017). Perancangan Basis Data Instrumen Bimbingan Konseling Alat Ungkap Pemahaman Diri Siswa. *Citec Journal*, 128-140.
- Juanita, S., & Wibisono, R. (2019). PEMODELAN SIA BIMBINGAN KONSELING DAN POIN PELANGGARAN SISWA BERBASIS DESKTOP STUDI KASUS SMK BINA BANGSA TANGERANG. *Jurnal MNEMONIC*, 5-11.
- Rana, S., & Garg, R. (2016). Application of Hierarchical Clustering Algorithm to Evaluate Students Performance of an Institute. *International Conference on Computational Intelligence & Communication Technology (CICT)*. Ghaziabad, India: IEEE.
- Triyanto, A. (2017). *Pengembangan Bahan/Materi Layanan Bimbingan dan Konseling berbasis Kebutuhan*. Sleman: CV Komojoyo Press.