

CLUSTERING DATA CALON SISWA BARU MENGGUNAKAN METODE K-MEANS DI SEKOLAH MENENGAH KEJURUAN WAHIDIN KOTA CIREBON

Asmana¹, Yudhistira Arie Wijaya², Martanto³

¹) Program Studi Teknik Informatika, STMIK IKMI Cirebon

²) Program Studi Sistem Informasi, STMIK IKMI Cirebon

³) Program Studi Manajemen Informatika, STMIK IKMI Cirebon

aaswijaya.id@gmail.com, yudhistira010471@gmail.com, martantomusijo@gmail.com

ABSTRAK

Pendidikan merupakan aspek yang sangat penting bagi kehidupan manusia, melalui pendidikan manusia dapat mempelajari bagaimana cara meningkatkan dan mengembangkan potensi yang di miliki, dunia pendidikan senantiasa mengalami perubahan yang lebih baik, berkembangnya teknologi menyebabkan masyarakat sangat membutuhkan informasi yang sangat cepat, akurat, instan dan efisien karena informasi merupakan bagian yang sangat penting untuk tetap menjalin komunikasi. Teknologi internet dapat memudahkan untuk mengakses data dan mencari berita maupun informasi secara online. Pemanfaatan teknologi informasi tidak hanya dimanfaatkan oleh perusahaan ataupun pemerintah. Namun penggunaan media online berbasis web ini digunakan dalam dunia pendidikan menjadi kebutuhan yang mendasar. Penggunaan web di sekolah tidak hanya digunakan untuk memberikan informasi online kepada civitas akademika dan komunitas, tetapi juga untuk mengoptimalkan website agar berfungsi sebagai sistem online yang mendukung sejumlah pekerjaan salah satunya adalah sistem informasi Penerimaan Peserta Didik Baru (PPDB).

Kata kunci: *clustering, K-means, siswa baru*

1. PENDAHULUAN

Algoritma K-Means merupakan algoritma clustering yang paling sederhana dibandingkan dengan algoritma lainnya. Algoritma ini memiliki kelebihan yaitu mudah diimplementasikan dan dijalankan, relatif cepat, mudah disesuaikan, dan paling banyak digunakan dalam tugas data mining. Algoritma ini merupakan salah satu algoritma terpenting dalam data mining. K-Means membagi data, mengelompokkannya menjadi beberapa cluster yang memiliki kesamaan, dan memisahkan setiap cluster. Hal ini didasarkan pada perbedaan antara setiap cluster. Algoritma ini diusulkan oleh sekitar peneliti di berbagai bidang [1]. Setiap sekolah memiliki data nilai siswa yang dapat diolah menggunakan teknologi data mining untuk menghasilkan pengetahuan yang bermanfaat dan bermanfaat. Salah satunya teknik data mining adalah clustering. Algoritma K-Means dapat digunakan untuk mengurutkan minat siswa dalam mata pelajaran matematika di sebuah sekolah sehingga dapat bermanfaat bagi mereka yang menggunakan kebijakan dalam pengambilan keputusan [2]. Algoritma clustering K-Means menjadi metode yang digunakan untuk menyelesaikan pengelompokan siswa sehingga diperoleh hasil yang dapat memetakan atau membagi kelompok siswa berdasarkan nilai siswa sebelumnya, Hal ini menjadi perhatian pendidikan untuk meningkatkan kualitas pembelajaran bagi siswa. Studi ini memberikan model pengelompokan bussines intelligence, metode yang digunakan adalah algoritma clustering K-Means dan algoritma greedy [3].

Penerimaan peserta didik baru merupakan kegiatan tahunan dari sekolah dan jumlah siswa baru

yang dapat mendaftar di SMK Taman Siswa Damai setiap tahun cenderung menurun. Oleh karena itu, diperlukan teknik untuk membantu mengubah data ini menjadi informasi yang berguna. Singkatnya, dengan menerapkan data penerimaan peserta didik baru sebagai strategi promosi dan menjadi dasar atau pedoman untuk strategi promosi berdasarkan wilayah di mana banyak siswa yang mendaftar. Penerapan metode klasterisasi K-Means untuk strategi promosi dapat membantu dalam proses pengelompokan data dalam bentuk hasil pengelompokan data [4]. Dengan bertambahnya jumlah sekolah dasar negeri dan swasta, ada banyak pilihan yang tersedia bagi orang tua sebelum mendaftarkan putra/putri mereka. Adanya permasalahan ketika orang tua ingin memilih sekolah dasar yang sesuai dengan kebutuhan dan kemampuan mereka dengan adanya clustering Algoritma K-Means membantu menentukan kriteria yang umumnya dijadikan dasar dalam pemilihan sekolah dasar [5]. MTs Darul Fikri Bringin selalu melihat peningkatan penerimaan siswa baru. Untuk menentukan kelas yang lebih tinggi, terdapat tingkat kesulitan yang mengelompokkan nilai sesuai dengan kemampuan siswa. Hal ini disebabkan oleh peningkatan jumlah pelanggan setiap tahunnya. Oleh karena itu, pada penelitian ini diterapkan metode data mining untuk membantu mengelompokkan kelas atas dengan poin siswa dengan menggunakan algoritma clustering K-Means dan tool Rstudio [6].

K-Means merupakan salah satu dari metode non-hierarchical clustering data yang dapat mengelompokkan data siswa menjadi cluster berdasarkan kemiripan data. Misalnya, data siswa dengan karakteristik yang sama dari akan

dikelompokkan menjadi satu cluster, dan data siswa dengan karakteristik yang berbeda dari akan menjadi Dikelompokkan ke dalam cluster lain yang dikelompokkan, atribut yang digunakan adalah asal sekolah, program studi dan nilai UAN, hasil penelitian ini akan menjadi dasar pengambilan keputusan dalam menentukan strategi untuk memajukan setiap program penelitian di Universitas Islam Indragiri [7]. SMK Syubbanul Wathon Magelang dalam melaksanakan langkah ini hanya dengan melihat nilai semester 1 sampai semester 5, sehingga proses mengelompokkan kemampuan siswa untuk memperkaya bahan ajar pengajaran di SMK SW masih kurang optimal. Untuk menambah keakuratan data maka diperlukan data nilai tambahan yaitu berupa tes tulis sesuai dengan kemampuan siswa Metode clustering K-Means digunakan untuk menyelesaikan masalah ini [8]. Data Mining atau disebut juga Knowledge Discovery in Database (KDD) adalah proses pengekstrasian pengetahuan dari data yang berukuran besar, mencari model baru untuk memperoleh pengetahuan dan informasi baru. Teknologi data mining digunakan untuk menemukan pengetahuan yang ada dalam database. Salah satu metode yang digunakan pada Data Mining adalah clustering dengan algoritma K-Means. Penelitian ini bertujuan untuk mengelompokkan siswa berusia untuk mencapai komposisi kelas yang seimbang yang meningkatkan hasil dan kualitas belajar siswa ketika nilai rata-rata di kelas dinaikkan dari [9].

Dalam proses penerimaan peserta didik baru PPDB di Sekolah Menengah Kejuruan Wahidin Kota Cirebon merupakan kegiatan awal sebelum dilakukannya kegiatan belajar mengajar, proses seleksi calon siswa dalam penerimaan siswa baru menjadi faktor penting dalam memberikan kemudahan pencapaian tujuan pendidikan yang diinginkan sekolah. Dalam hal ini SMK Wahidin Kota Cirebon masih sering menemui strategi yang tidak tepat dalam promosi penerimaan siswa dikarenakan tidak memiliki data induk dalam bentuk pengelompokan data asal sekolah siswa yang mendaftar di SMK Wahidin Kota Cirebon yang sudah dikaji secara ilmiah. Dengan ini kami merekomendasikan penggunaan algoritma k-means sebagai pengelompokan data asal sekolah siswa-siswi yang sudah mendaftar, yang sudah menjadi siswa-siswi yang sedang berjalan, dan juga alumni di SMK Wahidin untuk membantu memudahkan strategi promosi penerimaan peserta didik baru.

Pendidikan merupakan salah satu hal yang terpenting dalam kehidupan seseorang. Pendidikanlah yang menentukan dan menuntun masa depan dan arah hidup seseorang. Walaupun tidak semua orang berpendapat seperti itu, namun pendidikan tetaplah menjadi kebutuhan manusia. Bakat dan keahlian seseorang akan terbentuk dan terasah melalui pendidikan. Pendidikan juga sebagai tolak ukur kualitas seseorang. Berdasarkan permasalahan

sebelumnya pengelompokan data asal calon siswa baru kejuruan dan data asal calon peserta didik Sekolah Menengah Kejuruan Wahidin Kota Cirebon belum adanya informasi yang didapat berdasarkan bentuk tabel, grafik, atau prosesntasi penerimaan peserta didik baru dan berapa banyak jumlah asal sekolah yang mendaftarkan atau melanjutkan ke jenjang pendidikan Sekolah Menengah Kejuruan. Dengan adanya Algoritma K-means bisa digunakan untuk melakukan pengelompokan/clustering suatu data.

2. TINJAUAN PUSTAKA

Pada penelitian yang dilakukan oleh Ari Sulistiyawati dan Eko Supriyanto yang berjudul “Implementasi Algoritma K-Means Clustering dalam penentuan siswa kelas unggulan” Hasil penelitian menunjukkan bahwa perhitungan menggunakan metode K-Means Dengan mengintegrasikan atau mengimplementasikannya dalam sistem informasi, algoritma pengelompokan K-Means menghasilkan klarifikasi yang efisien dari subgrup dan pemrosesan iteratif, memutar jarak pusat, menentukan titik cluster mana yang terbentuk, dan hemat lebih banyak waktu dengan subkelompok kelas yang lebih tinggi [10].

Penelitian dengan judul “Algoritma K-Means dalam pemilihan siswa berprestasi dan metode SAW untuk prediksi Penerima Beasiswa Berprestasi” yang dilakukan oleh Rini Sovia, Eka Praja Wiyata Mandala, Sitty Mardhiah pada tahun 2020, Dalam hal ini yang menjadi permasalahan adalah beasiswa yang seringkali kurang tepat sasaran dan keterbatasan dana beasiswa membuat banyak siswa berprestasi tidak dapat memperolehnya, sehingga siswa berprestasi kehilangan semangat untuk terus belajar. Hasil penelitian adalah dengan menggabungkan K-Means dan SAW dapat mempermudah proses pendataan siswa berprestasi dan prediksi pemenang beasiswa karena lebih cepat dan lebih fokus, lebih sentral [11].

Penelitian dengan judul “Penerapan Algoritma K-Means Dalam Penentuan Jurusan Siswa Di SMAN 05 Bombana” yang dilakukan oleh Muh. Sulkifly Said dan Yusti dilakukan pada tahun 2020. Masalah Keterbatasan umum dalam mengambil jurusan di SMAN 05 Bombana adalah sulitnya mengidentifikasi siswa yang memenuhi kriteria untuk masuk ke dari jurusan. Algoritma K Means dapat mengelompokkan (segment) data yang memiliki atribut dan memiliki data dalam jumlah yang besar, sehingga dapat digunakandalam suatu sistem untuk menentukan jurusan siswa berdasarkan kemampuan belajar siswa [12].

Pada penelitian yang dilakukan oleh Dionisius Missa, Sentot Achmadi, Ali Mahmudi yang berjudul “Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Data Penghasilan Orang Tua Siswa” dilakukan pada tahun 2021. Untuk penelitian ini dibangun sebuah sistem yang dapat menerapkan metode K-Means, yang membantu

sekolah untuk menentukan klasifikasi data siswa yang memerlukan pembayaran ditanggung pada saat periode kenaikan sekolah dengan tujuan yang baik [13].

Penelitian dengan judul “Penerapan Algoritma K-Means Clustering Untuk Mengelompokkan Kenakalan Siswa” yang dilakukan oleh Harliana pada tahun 2020 ialah Tindak pidana anak di SMK Bina Insani dapat diklasifikasikan menjadi delik umum dan delik khusus. Di antara berbagai jenis pelanggaran tersebut, penelitian ini bertujuan untuk membantu guru BK mengelompokkan siswa yang menunggak menggunakan metode Kmeans berdasarkan jumlah pelanggarannya [14].

Berdasarkan penelitian yang dilakukan oleh Reynaldo Ferdian Saputra, Yosep Agus Pranoto, Renaldi Primaswara Prasetya yang di lakukan pada tahun 2021 dengan judul “Implementasi Metode K-Means Clustering Pada Tes Psikologi Berbasis Mobile Untuk Menentukan Kelompok Belajar Siswa” Metode yang digunakan adalah KMeans Clustering untuk menentukan hasil tes ditinjau dari kelompok gaya belajar siswa seperti auditori, visual, dan kinestetik [15].

Pada penelitian yang dilakukan oleh Muhammad Aliyul Amri, Dedy Hartama, Agus Perdana Windarto dengan judul penelitian “Penerapan Data Mining Pada Presentasi Penerimaan Imunisasi Anak-Anak Menggunakan K-Means Clustering” Dilakukan pada tahun 2020, penelitian ini membahas tentang penerapan data mining untuk mengelompokkan anak yang divaksinasi menurut provinsi menggunakan metode clustering Kmeans [16]

3. METODOLOGI

Pada penelitian ini penulis menggunakan metode penelitian yang digunakan adalah eksperimen. Metode yang bertujuan untuk menguji pengaruh suatu variabel terhadap variabel lain atau menguji bagaimana hubungan sebab akibat antara variabel yang satu dengan variabel yang lainnya. Metode penelitian eksperimen jelas berbeda dengan metode penelitian lainnya. Penelitian eksperimen dilakukan dengan membandingkan satu atau lebih kelompok eksperimen yang diberi perlakuan dengan satu atau lebih kelompok pembanding yang tidak diberi perlakuan.

3.1. Operasionalisasi Variabel

Pada pembahasan ini, variabel merupakan entitas/kesatuan apapun, yang mempunyai nilai yang berbeda/bervariasi. Pada penelitian dengan “Clustering Asal Calon Siswa Baru Menggunakan Metode K-Means Di Sekolah Menengah Kejuruan Wahidin Kota Cirebon memiliki 2 variabel yaitu:

- Variabel Bebas atau Variabel Independen (X): Algoritma K-means
- Variabel Tidak Bebas atau Variabel Dependen (Y): data asal calon siswa

Tabel 1. Operasional variabel

Variabel	Definisi / Data Variabel	Indikator (didukung oleh landasan teori)
Clustering data asal calon siswa (X)	Penerapan Algoritma K-Means Clustering dan Tools Rstudio dapat mengklaster penentuan dan penyeleksian kelas unggulan [17]	1. Data siswa 2. K-Means Clustering
K-Means	Data siswa asal sekolah harus dimiliki di setiap sekolah tanpa adanya data, sekolah kesulitan dalam mengolah data siswa [18]	1. Data Siswa 2. Data orang tua 3. Asal SMP

3.2. Sumber Data

Data primer merupakan data yang dapat diperoleh dari secara langsung ke bagian objek yang diteliti atau dari sumber orangnya, data tersebut berdasarkan data penerimaan peserta didik baru pada sekolah menengah kejuruan wahidin kota cirebon tahun pelajaran 2020/2021 kelas X yang melanjutkan ke jenjang sekolah menengah kejuruan. Data yang dilakukan bersumber dari website resmi SMK Wahidin Kota Cirebon dengan alamat <https://www.smkwahidincrb.sch.id/>

3.3. Teknik Pengumpulan Data

Teknik pengambilan data merupakan suatu cara yang dilakukan oleh peneliti untuk memperoleh data-data yang diperlukan. Dalam penelitian ini teknik yang digunakan antara lain sebagai berikut:

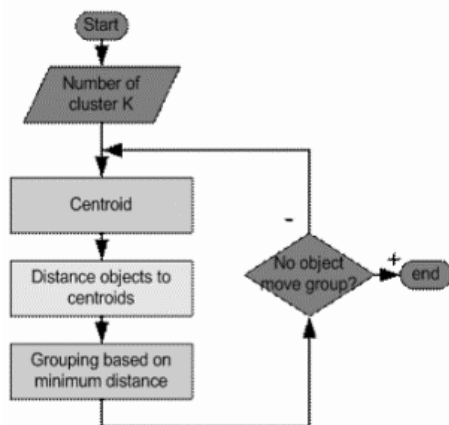
1. Wawancara
Wawancara dilakukan pada ketua Penerimaan Peserta Didik Baru yaitu dengan wakil kepala sekolah bagian kesiswaan dengan tujuan dapat mendalami permasalahan secara keseluruhan sehingga penelitian ini memiliki konsep untuk memecahkan masalah.
2. Survey
Survey penelitian dilakukan dengan meminta ijin kepada kepala sekolah SMK Wahidin Cirebon untuk pengambilan data mengenai data penerimaan data peserta didik baru dari tahun ajaran 2020/2021 yang akan dijadikan sebagai dataset dalam rapidminer.

3.4. Populasi dan Tempat

Populasi berjumlah 774 siswa yang diterima di Sekolah Menengah Kejuruan Wahidin Kota Cirebon. Berdasarkan data yang di terima pada tahun pelajaran 2020/2021. Tempat yang dijadikan sebagai objek penelitian berada di SMK Wahidin Kota Cirebon Jl. Dr. wahidin S (blkg. PG. Rajawali II NO. 21) Kec. Kejaksaan, Kota Cirebon, Jawa Barat 45122.

3.5. Metode Analisa Data

Pada penelitian ini menggunakan metode analisa Algoritma K-means clustering. K-Means clustering merupakan salah satu algoritma untuk menentukan klasifikasi objek berdasarkan atribut/karakteristik objek pada K cluster/partisi. K adalah bilangan positif yang menunjukkan banyaknya kelompok/cluster/partisi dari objek tersebut. Pemisahan data dilakukan dengan mencari nilai jarak minimum antara data dengan nilai centroid. Ini ditentukan secara acak atau menggunakan set centroid awal. Anda juga dapat menentukan nilai centroid berdasarkan k objek yang berurutan. Untuk tahapan detailnya tentang K-Means bisa dilihat pada gambar dan pembahasan dibawah ini :



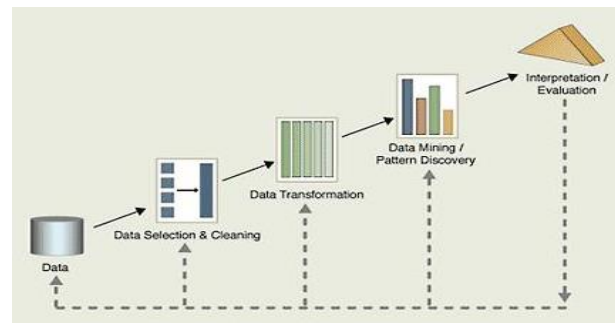
Gambar 1. Penerapan K-Means Cluster

Centroid adalah rata-rata aritmatika bentuk objek dari setiap titik dalam objek. Penerapan klasterisasi K-Means dapat dilakukan dengan prosedur langkah demi langkah berikut.

- a) Menyiapkan data training berbentuk vector
- b) Siapkan nilai K cluster
- c) Set nilai awal centroids.
- d) Gunakan rumus jarak Euclidean untuk menghitung jarak antara data dan centroid
- e) Partisi data berdasarkan nilai minimum
- f) Kemudian ulangi saat partisi data masih bergerak (tidak perlu memindahkan objek ke partisi lain). Jika Anda masih belum, lanjutkan ke poin 3
- g) Jika dataset saat ini sama dengan dataset sebelumnya, hentikan iterasi.
- h) Data dibagi sesuai dengan nilai centroid akhir.

3.6. Tahapan Penelitian

Proses menganalisis data dalam penerapan data mining menggunakan tahapan knowledge discovery in databases (KDD) yang terdiri dari Data, Data Cleaning, Data transformation, Data mining, Pattern evolution, knowledge:



Gambar 2. Tahapan KDD

Dibawah ini adalah pembahasan tentang tahap knowledge discovery in databases (KDD)

1. Data Selection
Data PPDB pada SMK Wahidin Kota Cirebon diambil pada tahun pelajaran 2020/2021 Data yang digunakan dalam penelitian ini sebanyak 774 record
2. Data Cleaning
Pada umumnya data yang diperoleh, baik dari database PPDB SMK Wahidin Kota Cirebon maupun survey, terdapat isian-isian yang tidak lengkap seperti data hilang, data tidak valid, dan salah ketik. Lebih baik membuang data yang tidak relevan, karena keberadaannya nantinya dapat mengurangi kualitas dan akurasi hasil data mining. Pembersihan data juga berdampak pada kinerja dari system data mining karena mengurangi jumlah dan kompleksitas data yang diproses.
3. Data integration
Integrasi data dilakukan pada atribut yang mengidentifikasi entitas unik SMK Wahidin Cirebon dengan 774 atribut Integrasi data harus dilakukan dengan hati-hati, karena kesalahan integrasi data dapat menyebabkan hasil data yang bias dan dapat menyesatkan tindakan yang diambil kemudian.
4. Data transformation
Merupakan proses transformasi pada data terpilih, sehingga data tersebut sesuai untuk proses data mining.
5. Data Mining
Data mining adalah proses menemukan dan menganalisis data dalam jumlah besar dengan tujuan menemukan pola atau informasi yang menarik dari sejumlah besar data yang tersimpan dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma yang tepat sangat bergantung pada tujuan dan proses knowledge discovery in databases (KDD) secara keseluruhan. Tahap ini merupakan inti dari tahapan knowledge discovery in databases (KDD) yang dilakukan untuk menganalisis data yang telah dibersihkan.
6. Pattern evaluation
Pada langkah ini, hasil dari teknik data mining berupa pola dan model dievaluasi untuk menentukan apakah benar-benar dapat dicapai.

Jika hasil yang diterima tidak konsisten, beberapa alternatif dapat diterapkan, seperti memberikan umpan balik untuk memperbaiki data mining lain yang lebih sesuai atau menerima hasil sebagai hasil yang tidak sesuai harapan dapat membantu.

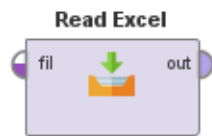
7. Knowledge

Langkah terakhir dalam proses data mining adalah membentuk keputusan atau tindakan berdasarkan hasil analisis.

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Tahapan pertama adalah mencari file lokasi yang sudah dibuat sebelumnya dengan format *xlsx* menggunakan Operator Read Excel yang berfungsi membaca data *xlsx*. Pada gambar 3 menampilkan data yang sudah di *import*, sehingga tombol pada *operator read excel* tidak lagi tanda seru warna kuning yang artinya operator telah berisi data dan siap diolah.



Gambar 3. Operator Read Excel

NAMA LENGKAP	NIS	KELAS	Jeni Kelamin	Tempat Lahir	Tanggal
polynomial	integer	polynomial	polynomial	polynomial	integer
1 ANINDYA RIZTA...	202110288	X AKL 1	Perempuan	CIREBON	30
2 ALIZZA	202110289	X AKL 1	Perempuan	CIREBON	3
3 ANGGUN SOPRY...	202110291	X AKL 1	Perempuan	CIREBON	10
4 APRIYANTO ARD...	202110292	X AKL 1	Laki-laki	CIREBON	27
5 ARSY JANNAH A...	202110293	X AKL 1	Perempuan	CIREBON	25
6 ASTI RAHMAWATI	202110294	X AKL 1	Perempuan	CIREBON	23
7 AULIYA INSANI	202110295	X AKL 1	Perempuan	CIREBON	11
8 BANYU BENING	202110296	X AKL 1	Perempuan	CIREBON	6
9 CUT NABILA ZAH...	202110297	X AKL 1	Perempuan	CIREBON	6
10 DEDE IKHSAN	202110298	X AKL 1	Laki-laki	KUNINGAN	4
11 DELA MONIKA	202110299	X AKL 1	Perempuan	CIREBON	7
12 DINA SAFIRA	202110300	X AKL 1	Perempuan	CIREBON	26

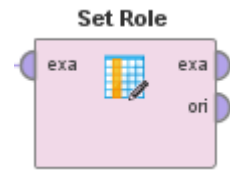
Gambar 4. Data dari excel

Selanjutnya menggunakan operator set role berfungsi buat membedakan baris penamaan atribut

Tabel 3. Sebelum dilakukan Preprocessing

1	NO	NAMA LENGKAP	NIS	KELAS	Jeni Kelamin	Tempat Lahir	Tanggal	Bulan	Tahun	Asal SMP/MTS	Nama Desa/Kelurahan	Kecamatan	Kota/Kabupaten	Nama Ayah Kandung	Nama Ibu Kandung
2	1	AININDYA RIZTAFIANI	202110288	X AKL 1	Perempuan	CIREBON	30	JULI	2005	SMP N 10 CIREBON	Kesambi	Kesambi	Kota Cirebon	ADI FEBIARDI	TRIA JAMIATI
3	2	ALIZZA	202110289	X AKL 1	Perempuan	CIREBON	3	DESEMBER	2004	SMP N 12 CIREBON	Ciperna	Talun	Kab. Cirebon	MAIMAN	FITRIANAH
4	3	ANGGUN SOPRIYANA	202110291	X AKL 1	Perempuan	CIREBON	10	DESEMBER	2005	SMP N 2 SURANENGGALA	Suranenggala Kidul	Suranenggala	Kab. Cirebon	SUHANDI	CAWI
5	4	APRIYANTO ARDI PRABOWO	202110292	X AKL 1	Laki-laki	CIREBON	27	APRIL	2004	SMP N 4 CIREBON	Kesambi	Kesambi	Kota Cirebon	ARDIAN	SUHERLINA INDAH
6	5	ARSY JANNAH ALMUBAROKAH	202110293	X AKL 1	Perempuan	CIREBON	25	JANUARI	2005	SMP N 3 CIREBON	Pegambiran	Lemahwungkuk	Kota Cirebon	SLAMET HARTONO	YUSNIAWATI
7	6	ASTI RAHMAWATI	202110294	X AKL 1	Perempuan	CIREBON	23	NOVEMBER	2004	SMP N 9 CIREBON	Hajjamukti	Hajjamukti	Kota Cirebon	MOHAMMAD ROKHATA	YULIANA
8	7	AULIYA INSANI	202110295	X AKL 1	Perempuan	CIREBON	11	MARET	2005	SMP N 2 SURANENGGALA	Surakarta	Suranenggala	Kab. Cirebon	DUGANI	AYATUN LUPUS
9	8	BANYU BENING	202110296	X AKL 1	Perempuan	CIREBON	6	MEI	2005	SMP N 8 CIREBON	Hajjamukti	Hajjamukti	Kota Cirebon	TRIS SATRIANA	YUYUN RAMADIAH
10	9	CUT NABILA ZAHRA	202110297	X AKL 1	Perempuan	CIREBON	6	FEBRUARI	2005	SMP N 14 CIREBON	Pasindangan	Gunung Jati	Kab. Cirebon	FERRIYANTO	SUKESIH
11	10	DEDE IKHSAN	202110298	X AKL 1	Laki-laki	KUNINGAN	4	DESEMBER	2004	SMP N 7 CIREBON	Kesambi	Kesambi	Kota Cirebon	ERWANTO	NUNUNG ROHANA
12	11	DELA MONIKA	202110299	X AKL 1	Perempuan	CIREBON	7	DESEMBER	2004	SMP N 14 CIREBON	Lemahwungkuk	Lemahwungkuk	Kota Cirebon	MAKADI	SUHAIRYANI
13	12	DINA SAFIRA	202110300	X AKL 1	Perempuan	CIREBON	25	AGUSTUS	2005	MTS KAPETAKAN	Kertapura	Kapetakan	Kab. Cirebon	SUNAJI	RUBINA
14	13	DINI PRAWATI	202110301	X AKL 1	Perempuan	CIREBON	25	NOVEMBER	2005	SMP N 13 CIREBON	Pegambiran	Lemahwungkuk	Kota Cirebon	SUMARDI	SITI ROKAYAH
15	14	EKA PUTRA PRATAMA	202110302	X AKL 1	Laki-laki	CIREBON	30	MARET	2005	SMP N 2 CIREBON	Kejaksan	Kejaksan	Kota Cirebon	KAMALUDIN	ENDANG MARIYATI
16	15	FAHRI SAPUTRA	202110303	X AKL 1	Laki-laki	CIREBON	19	JULI	2006	SMP N 3 GUNUNG JATI	Sambeng	Gunung Jati	Kab. Cirebon	ANDI	YANI
17	16	FANI FEBRIYANI	202110304	X AKL 1	Perempuan	CIREBON	22	DESEMBER	2004	SMP N 1 GUNUNG JATI	Sambeng	Gunung Jati	Kab. Cirebon	DEDI SURRIYANTO	SRI MULYA
18	17	FARHAN DWI AGUSTIAN	202110305	X AKL 1	Laki-laki	CIREBON	3	AGUSTUS	2005	SMP WAHIDIN CIREBON	Sukapura	Kejaksan	Kota Cirebon	ANDI SURACHMAN	NOVA LISTIANTI
19	18	FASHI HULLISAN	202110306	X AKL 1	Laki-laki	CIREBON	20	JUNI	2005	SMP N 1 SURANENGGALA	Karangreja	Suranenggala	Kab. Cirebon	BASARUDIN	MAUNA
20	19	FITRI OKTAVIANI	202110307	X AKL 1	Perempuan	CIREBON	30	OKTOBER	2004	MTS N 4 CIREBON	Tegalwangi	Weru	Kab. Cirebon	MOHAMMAD ANWAR	YUHANAH
21	20	JESICA AFRISCA	202110308	X AKL 1	Perempuan	CIREBON	7	APRIL	2005	SMP N 14 CIREBON	Kesenden	Kejaksan	Kota Cirebon	EDI JUNAEDI	YANI MULYANI

koordinat & prediksi posisi yg akan pada masukan kedalam kategori 'label'. Agar dalam ketika pengkategorian data 'label' dan terhitung & merubah hasil.



Gambar 5. Set Role

Tabel 2. Parameter set role

Parameter	Nilai	
attribute name	NIS	
target role	Id	
set additional roles	attribute name	target role
	NIS	Id

Gambar 6. Hasil Set Role

4.2. Preprocessing

Pada tahapan preprocessing ini bertujuan untuk menghilangkan missing value pada data yang akan digunakan sehingga pada saat diproses tidak muncul kesalahan atau masalah. Berikut table sebelum dilakukan preprocessing se-bagai berikut.

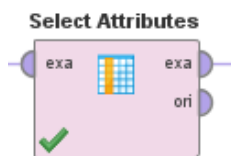
Preprocessing data dilakukan dengan menggunakan software RapidMiner versi 9.10. Setelah dilakukan preprocessing data, 7.5 cm didapatkan hasil bahwa tidak ada missing value pada data, baik pada atribut maupun label.

Tabel 4. Sesudah Preprocessing

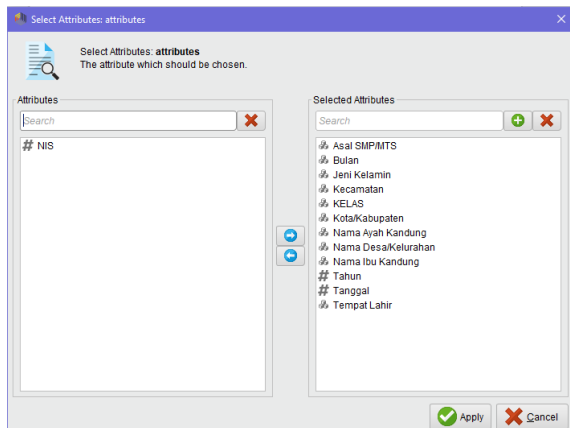
ID	NIS	Nilai	0	17187027	202111785	20200906873
KELAS	Pideneria	0	X ODP 4 (31)	X BOP 4 (38)	X BOP 4 (38)	X INU 1 (38) ... (38 min)
Jeni Kelamin	Pideneria	0	Laki-laki (254)	Pemempuan (462)	Pemempuan (462)	Pemempuan (462)
Tempat Lahir	Pideneria	0	WOREBOBO (1)	CIREBON (723)	CIREBON (723)	WHERES (8) ... (8 min)
Tanggal	Integer	0	1	31	10.286	
Bulan	Pideneria	0	FEBRUARI (31)	APRIL (30)	APRIL (30)	JUNI (30) ... (30 min)
Tahun	Integer	0	2002	2004	2004.190	
Asal SMP/MTS	Pideneria	0	SMRT PEE JATI (1)	SMP WAHIDIN CIREBON	SMP WAHIDIN CIREBON (7)	SMPN 2 GURUG JATI (40) ... (40 min)
Nama Desa/Kelurahan	Pideneria	0	Setatikan (1)	Pegajaran (35)	Pegajaran (35)	Sekeloa (1) ... (1 min)
Kecamatan	Pideneria	0	Tayang Prih (1)	Demung Jati (140)	Demung Jati (140)	Lembahwang (87) ... (87 min)
Kota/Kabupaten	Pideneria	0	Kota Cirebon (1)	Kab. Cirebon (470)	Kab. Cirebon (470)	Kota Cirebon (1) ... (1 min)
Nama Ayah Kandung	Pideneria	0	ZAWARA (1)	BILAEWAY (5)	BILAEWAY (5)	DESI (1) ... (1 min)
Nama Ibu Kandung	Pideneria	0	YULIAN YUHEL (1)	NURWATI (7)	NURWATI (7)	USARIN (4) ... (4 min)

Selanjutnya menggunakan operator Select Attributes. Operator ini menyediakan berbagai jenis filter untuk memudahkan pemilihan Atribut. Kemungkinannya misalnya. Pemilihan langsung Atribut. Seleksi dengan ekspresi reguler atau hanya memilih Atribut tanpa nilai yang hilang. Hanya Atribut yang dipilih yang dikirim ke port output. Sisanya dihapus dari ExampleSet.

Di bawah ini adalah gambar 7 menunjukkan hasil dari Select Attributes.



Gambar 7. Select Attributes

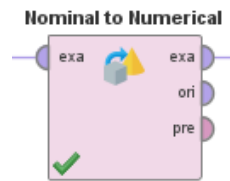


Gambar 8. Hasil Select Attributes

4.3. Transformation

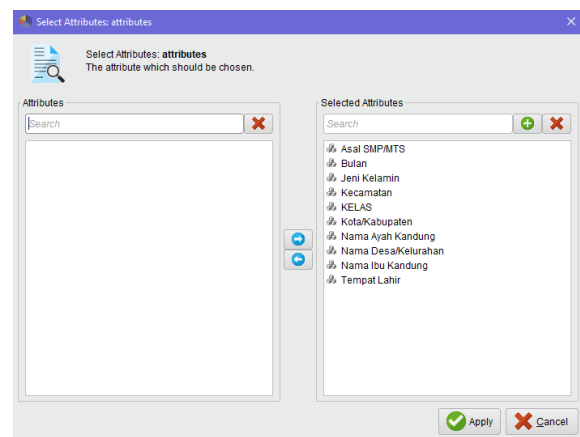
Pada tahap transformation ini peneliti mencantumkan transformation dikarenakan mengubah tipe atribut non-numerik yang dipilih

menjadi tipe numerik. Ini juga memetakan semua nilai atribut ini ke nilai numerik. Untuk algoritma yang di buat menggunakan algoritma K-Means.



Gambar 9. Nominal to Numerical

Untuk melihat hasil dari nominal to numerical bisa dilihat pada gambar 10 Hasil Nominal to Numerical



Gambar 10. Hasil Nominal to Numerical

4.4. Data Mining

Penentuan jumlah cluster merupakan langkah yang harus dilakukan. Nilai kluster atau k yang digunakan adalah (jumlah cluster optimal, jumlah anggota optimal dalam cluster). Hasil dari pengelompokan tersebut dimana akan dijadikan sebagai acuan untuk melakukan penerimaan data asal sekolah Sekolah Menengah Kejuruan (SMK) Wahidin Kota Cirebon.

Berikut adalah desain proses untuk membuat cluster K-Means data asal calon siswa baru di SMK WAHIDIN KOTA CIREBON menggunakan tool Rapid Miner versi 11 :



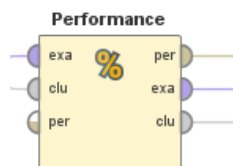
Gambar 11. Desain K-Means

Pada gambar 12 adalah operator utama Clustering K-Means



Gambar 12. Cluster K-Means

Clustering merupakan operator Cluster K-Means, operator ini adalah operator utama dari pemodelan proses ini agar dapat mengetahui hasil dari pengklasteran tersebut. Algoritma K-Means menentukan satu set k cluster dan menetapkan setiap Contoh ke satu cluster yang tepat.



Gambar 13. Performance

Performance di gunakan untuk mengukur Cluster Distance Performance dari proses K-Means tersebut, untuk mengukur peformaa dari operator K-Means. Operator ini digunakan untuk evaluasi kinerja metode clustering berbasis centroid. Operator ini memberikan daftar nilai kriteria kinerja berdasarkan centroid cluster. ni juga memiliki informasi mengenai centroid dari setiap cluster. Operator Cluster Distance Performance mengambil model cluster centroid ini dan clustered set sebagai input dan mengevaluasi kinerja model berdasarkan centroid cluster. Dua ukuran kinerja didukung: Rata-rata dalam jarak cluster dan indeks Davies-Bouldin. Ukuran kinerja ini dijelaskan dalam parameter.

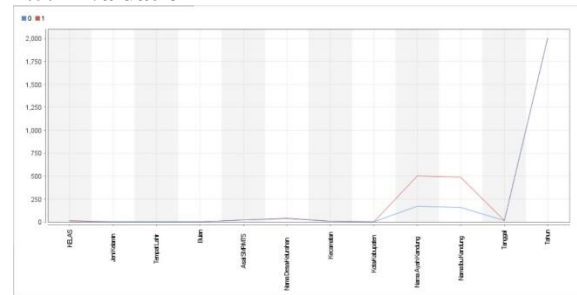
PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -28328.241
Avg. within centroid distance_cluster_0: -23694.058
Avg. within centroid distance_cluster_1: -32751.780
Davies Bouldin: -0.619
```

Gambar 14. Hasil Performance Vector

Hasil dari pengelompokan menggunakan algoritma K-Means dengan menggunakan tool Rapid Miner versi 9.10 untuk mencari K2 - K20 memperoleh hasil K mana yang paling baik performanya yaitu yang mendekati angka 0. K yang mendekati angka 0 adalah K2. Dari K2=0,619 dan iterasi ke 1 dari jumlah iterasi 1 - 50, untuk mengetahui peforma dari algoritma K-Means dapat dilihat dari performance. Dimana davies bouldin apabila mendekati angka 0 nilai berarti algoritma semakin baik, maka jumlah cluster terbaik disini ada 2 yang terbaik atau yang mendekati 0.

4.5. Evaluation



Gambar 15. Hasil Evaluation

Maka hasil dari clustering data asal sekolah dengan Algoritma K-Means diperoleh dari kelas, jenis kelamin, tempat lahir, bulan, asal SMP, nama kelurahan atau desa, kecamatan memperoleh hasil yang sama bisa dilihat pada gambar 5.4 hasil Evaluation dan nilai Support tertinggi ada pada tanggal lahir siswa dengan akurasi 2,000 terdapat tanggal lahir yang berbeda-beda di setiap siswanya. Nama ibu kandung dan bapak kandung memperoleh hasil 500, dan kota atau kabupaten.

4.6. Pembahasan

Pada penelitian yang dilakukan oleh Wijaya [19] tentang Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities mendapat hasil *Davies Bouldin Index* terbaiknya 0,168 measure types menggunakan operator *Mixed Measures* dan nilai k = 2 (0,168), k = 3 (0,248), k = 4 (0,232), k = 5 (0,26), k = 6 (0,197), pada penelitian yang dilakukan peneliti menggunakan measure types dengan 2 operator yang berbeda yaitu operator *Numerical Measures* dan Operator *Bregman Divergences* dan dihasilkan *Davies Bouldin Index* pada yang sama yaitu 0,619 nilai terbaiknya dari kedua operator tersebut

5. KESIMPULAN DAN SARAN

Maka berdasarkan penelitian yang telah dilakukan dapat disimpulkan bawah Algoritma K-Means dapat mengalisis clustering data asal calon siswa mencari informasi tentang penerimaan peserta didik baru di SMK Wahidin Kota Cirebon sebanyak 774 siswa

DAFTAR PUSTAKA

- [1] Tandy, J., & Assegaff, S. (2019). Analisis Dan Perancangan Clustering Siswa Baru Menggunakan Metode K-Means. *Manajemen Sistem Informasi*, 4(4), 389–399.
- [2] Butsianto, S., & Saepudin, N. (2020). Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 3(1), 51–59. <https://doi.org/10.32672/jnkti.v3i1.2008>
- [3] Amirulloh, I. (2019). Pemetaan Kelompok Kerja Siswa Dengan Metode Clustering K-Means Dan Algoritma Greedy. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 1(2), 94–98. <https://doi.org/10.36499/jinrpl.v1i2.2953>
- [4] Purwanti, T., Ramdhan, W., & Santoso, S. (2021). Penerapan Metode Klasterisasi K-Means untuk Strategi Promosi Pada SMK Tamansiswa Sukadamai. *JUTSI (Jurnal Teknologi Dan Sistem Informasi)*, 2(1), 141–146. <https://doi.org/10.33330/jutsi.v2i1.1156>
- [5] Sari, D. W., & Ayu, K. G. (2019). Penentuan Kriteria Dalam Memilih Sekolah Dasar Dengan Menerapkan K-Means Clustering (Studi Kasus: Wilayah Kecamatan Mampang). *Sainstech: Jurnal Penelitian Dan Pengkajian Sains Dan Teknologi*, 29(2), 24–28. <https://doi.org/10.37277/stch.v29i2.334>
- [6] Primanda, R. P., Alwi, A., & Mustikasari, D. (2021). DATA MINING SELEKSI SISWA BERPRESTASI UNTUK MENENTUKAN KELAS UNGGULAN MENGGUNAKAN METODE K-MEANS CLUSTERING (Studi Kasus di MTS Darul Fikri). *Komputek*, 5(1), 88. <https://doi.org/10.24269/jkt.v5i1.686>
- [7] Yunita, F. (2018). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru. *Sistemasi*, 7(3), 238. <https://doi.org/10.32520/stmsi.v7i3.388>
- [8] Hartanti, N. T., Informasi, S., Komputer, I., Indonesia, B., Inggris, B., Inggris, B., & Indonesia, B. U. R. (2018). Education Data Mining Untuk Menentukan Kelompok. *Jurnal Tekno Kompak*, 12(2), 39–44.
- [9] Elda, Y., Defit, S., Yunus, Y., & Syaljumairi, R. (2021). Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means. *Jurnal Informasi Dan Teknologi*, 3, 103–108. <https://doi.org/10.37034/jidt.v3i3.130>
- [10] Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*, 15(2), 25. <https://doi.org/10.33365/jtk.v15i2.1162>
- [11] Sovia, R., Mandala, E. P. W., & Mardhiah, S. (2020). Algoritma K-Means dalam Pemilihan Siswa Berprestasi dan Metode SAW untuk Prediksi Penerima Beasiswa Berprestasi. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 6(2), 181. <https://doi.org/10.26418/jp.v6i2.37759>
- [12] Said, M. S., & Yusti, Y. (2020). Penerapan Algoritma K-Means Dalam Penentuan Jurusan Siswa Sman 05 Bombana. *Simtek: Jurnal Sistem Informasi Dan Teknik Komputer*, 5(2), 114–122. <https://doi.org/10.51876/simtek.v5i2.87>
- [13] Missa, D., Achmadi, S., & Mahmudi, A. (2021). Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Data Penghasilan Orang Tua Siswa. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(1), 125–133. <https://doi.org/10.36040/jati.v5i1.3275>
- [14] Harliana, H. (2020). Penerapan Algoritma K-Means Clustering Untuk Mengelompokkan Kenakalan Siswa. *Journal of Information System Management (JOISM)*, 2(1), 19–22. <https://doi.org/10.24076/joism.2020v2i1.200>
- [15] Saputra, R. F., Pranoto, Y. A., & Industri, F. T. (2021). IMPLEMENTASI METODE K-MEANS CLUSTERING PADA TES PSIKOLOGI. 5(1), 328–333.
- [16] Amri, M. A., Hartama, D., & Windarto, Ap. (2020). Penerapan Data Mining Pada Presentasi Penerimaan Imunisasi Anak-Anak Menurut Provinsi Menggunakan K-Means Clustering. *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 319–324.
- [17] Oktarian, S., Defit, S., & Sumijan. (2020). Clustering Students' Interest Determination in School Selection Using the K-Means Clustering Algorithm Method. *Jurnal Informasi Dan Teknologi*, 2, 68–75. <https://doi.org/10.37034/jidt.v2i3.65>
- [18] Nur, F., Zarlis, M., & Nasution, B. B. (2017). Penerapan Algoritma K-Means Pada Siswa Baru Sekolahmenengah Kejuruan Untuk Clustering Jurusan. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, 1(2), 100–105. <https://doi.org/10.30743/infotekjar.v1i2.70>
- [19] Wijaya, Y. A., Kurniady, D. A., Setyanto, E., Tarihoran, W. S., Rusmana, D., & Rahim, R. (2021). Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities. *TEM Journal*, 10(3), 1099–1103. <https://doi.org/10.18421/TEM103-13>