

ANALISA DATA SET PEMINATAN SISWA MENGGUNAKAN ALGORITMA K-MEANS DENGAN OPTIMIZE PARAMETER DI SEKOLAH MENENGAH KEJURUAN (STUDI KASUS: SMK PUI GEGESIK)

Rahmat Sidik, Nana Suarna, Arif Rinaldi Dikananda

Teknik Informatika, STIMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Indonesia

Mamatsenpai69@gmail.com

ABSTRAK

Banyaknya siswa yang ragu akan minat karir adalah hal yang sudah biasa terjadi di SMK PUI Gegesik, maka dari itu SMK PUI Gegesik membuat 3 program untuk para siswanya yaitu Campus camp yang berfokus minat siswa yang ingin melanjutkan ke jenjang perguruan tinggi, Program BKK yang berfokus ke pelatihan siswa menghadapi dunia kerja, dan Wirausaha muda yang di khususkan berfokus untuk menyiapkan para siswa untuk berwirausaha di usia muda. Dengan ini penulis merekomendasikan menggunakan algoritma K-means untuk pengelompokan minat siswa berdasarkan data minat siswa dan menggunakan Optimize parameter untuk menemukan nilai optimal dari parameter yang dipilih untuk Operator dalam subprosesnya. Tujuan dari penelitian ini adalah untuk mencari parameter yang paling baik dalam pengelompokan data set minat siswa menggunakan Algoritma K-means untuk menjadikan data dan informasi yang diperoleh untuk pihak SMK PUI Gegesik. Pada penelitian ini metode eksperimen digunakan untuk mengelompokkan minat siswa dan mengetahui parameter mana yang paling baik digunakan oleh algoritma K-Means. Dengan melakukan uji beda dengan jumlah cluster ($k = 2, 3, 4, 5, 6, 7, 8, 9, 10$), hasil metode k-means dengan tipe measure adalah Mixed Measures dengan DBI paling optimal sebesar 0,497 dengan 2 cluster.

Kata kunci: Minat, K-means, Clustering, Data Mining

1. PENDAHULUAN

Data Mining adalah proses menemukan atau mengekstraksi informasi yang terjadi dalam data besar. Clustering adalah suatu metode yang digunakan dalam proses Data Mining. Dalam clustering terdapat beberapa metode, K-means adalah salah satunya. K-means adalah proses metode clustering non-hierarchical yang membagi data menjadi satu atau lebih cluster [1]. Proses penentuan jurusan bagi siswa SMK berulang setiap tahunnya. Dengan jumlah data yang begitu besar, mudah terjadi kesalahan dalam proses pemilihan minat jurusan. Algoritma K-means clustering pada penelitian ini dibatasi penerapannya menggunakan tools data mining berupa RapidMiner Studio. Metode yang digunakan adalah metode clustering. Clustering adalah pengelompokan record, observasi, atau instance ke dalam kelas fitur yang serupa. Hasil penerapan algoritma K-means clustering pada penelitian ini telah memberikan solusi yang baik untuk membagi magang untuk jurusan berdasarkan minat dan bakat siswa kelas 11 Komunikasi multimedia dan akuntansi [2].

Penelitian ini penting untuk diteliti karena selama ini permasalahan belum adanya penelitian tentang pengelompokan data menggunakan data minat siswa sehingga dibuatkan pengolahan data minat siswa dengan menggunakan metode K-means. Dalam penerapan K-Means, belum ada kepastian dalam menemukan akurasi cluster yang terbentuk. Oleh karena itu, diperlukan validasi dalam mengukur seberapa baik atau akurat klaster yang dihasilkan. Davies Bouldin Index (DBI) merupakan salah satu

metode validitas dalam clustering. Pendekatan DBI memaksimalkan jarak antar-cluster dan meminimalkan jarak intra-cluster [3].

Penelitian terdahulu oleh Yudhistira Arie Wijaya, Penelitian ini tentang Algoritma Indeks Davies Bouldin untuk Mengoptimalkan Studi Kasus Clustering Pemetaan Fasilitas Sekolah. Indeks Davies Bouldin (DBI) yang lebih rendah dianggap sebagai algoritma pengelompokan terbaik berdasarkan kriteria yang menghasilkan kumpulan klaster. Tujuan dari penelitian ini adalah untuk mengoptimalkan hasil clustering menggunakan DBI. Sumber data yang digunakan adalah jumlah desa yang memiliki fasilitas sekolah dan tingkat pendidikan yang diperoleh dari website pemerintah (<https://www.bps.go.id>). Jenjang pendidikan yang dimaksud adalah SMA dan SMK. Metode yang digunakan adalah k-means. Hasil menunjukkan bahwa dari jumlah cluster ($k = 2, 3, 4, 5, 6$) diperoleh DBI optimal untuk ($k = 2$) dengan nilai 0,168 untuk Measure Type = Mixed Measures. Untuk nilai $k = 2$ diperoleh pemetaan wilayah dengan L0 (rendah) = 31 provinsi dan L1 (tinggi) = 3 provinsi. Centroid akhir yang diperoleh untuk setiap cluster adalah L0 (315 dan 155) dan L1 (1710 dan 1259) [4]. Kumpulan data tentang nilai siswa di suatu sekolah dapat digunakan dengan teknologi data mining untuk menghasilkan suatu pengetahuan yang menarik dan bermanfaat yang selama ini belum diketahui secara manual. Algoritma K-Means dapat dipakai untuk mengurutkan minat siswa terhadap matematika di sekolah, sehingga dapat bermanfaat bagi pengguna kebijakan dalam proses pengambilan keputusan [5].

Banyaknya siswa yang ragu akan minat karir adalah hal yang sudah biasa terjadi di SMK PUI Gegesik, maka dari itu SMK PUI Gegesik membuat 3 program untuk para siswanya yaitu Campus camp yang berfokus minat siswa yang ingin melanjutkan ke jenjang perguruan tinggi, Program BKK yang berfokus ke pelatihan siswa menghadapi dunia kerja, dan Wirausaha muda yang di khususnya berfokus untuk menyiapkan para siswa untuk berwirausaha di usia muda. Dalam hal ini SMK PUI Gegesik masih menemui strategi yang tidak tepat dalam pembekalan siswa dikarenakan tidak memiliki data minat siswa dalam bentuk pengelompokan data minat siswa yang sudah dikaji secara ilmiah. Dengan ini penulis merekomendasikan menggunakan algoritma K-means untuk pengelompokan minat siswa berdasarkan data minat siswa dan menggunakan Optimize parameter untuk menemukan nilai optimal dari parameter yang dipilih untuk Operator dalam subprocessnya.

Kajian yang di usulkan adalah yaitu menggunakan K-means dan operator Optimize parameter untuk mendapatkan hasil optimal dari sebuah paramter, dan pada penelitian ini menggunakan jenis pengukuran dengan tiga parameter yang berbeda yaitu parameter Mixed measure, numerical measure dan bregmen divergences, dan menghasilkan Davies Bouldin nilai index (dbi) terbaik dari sebuah data. Berdasarkan permasalahan sebelumnya SMK PUI Gegesik belum memiliki informasi berdasarkan table, grafik dan persentasi tentang minat para siswa.

Tujuan dari penelitian ini adalah untuk mencari parameter yang paling baik dalam pengelompokan data set minat siswa menggunakan Algoritma K-means untuk menjadikan data dan informasi yang diperoleh untuk pihak SMK PUI Gegesik sebagai acuan tolak ukur untuk siswa angkatan selanjutnya dan menjadikan data ini referensi untuk pengetahuan tentang minat siswa. Dalam pengelompoakan ini menggunakan optimize parameter untuk mendapatkan nilai yang optimal dari parameter.

Berdasarkan pemaparan yang penulis susun bermaksud mengangkat judul penelitian ini yaitu "Analisa Data Set Peminatan Siswa Menggunakan Algoritma K-means Dengan Optimize Parameter Di Sekolah Menengah Kejuruan SMK PUI Gegesik", alasan peneliti mengambil judul tersebut karena masih belum adanya informasi pengelompokan minat siswa dan penulis ingin menerapkan metode clustering menggunakan algoritma K-means untuk mengelompokan minat siswa yang dimana algoritma K-means ini efisien dalam melakukan klasterisasi serta menggunakan optimize parameter agar mendapat hasil lebih optimal.

2. TINJAUAN PUSTAKA

2.1. Data mining

Data mining adalah istilah yang digunakan untuk menggambarkan menemukan informasi dalam database. Data mining juga merupakan teknik yang

menggabungkan teknik analisis data dan menemukan pola dalam data [6]

2.2. Clustering

Clustering pada dataset merupakan langkah untuk mengelompokan *dataset* yang atribut kelasnya belum terdeskripsikan, *clustering* konseptual melibatkan pemaksimalan dan peminimalan kemiripan antar kelas [7]

2.3. K-means

Algoritma *K-means* adalah metode pengelompokan non-hierarkis yang mempartisi objek menjadi beberapa kelompok (*cluster*). Teknik ini bertujuan untuk meminimalkan variasi dalam suatu *cluster* dan memaksimalkan variabilitas antar *cluster* [8]

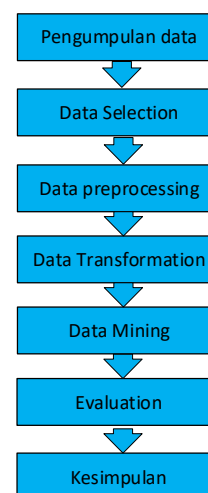
2.4. K-means Clustering

K-means clustering adalah salah satu teknik *clustering* dalam proses pemodelan *Data Mining* tanpa pengawasan dan metode pengelompokan data berdasarkan partisi. Data dikelompokkan menggunakan metode *K-means* menjadi beberapa kelompok dan setiap kelompok mempunyai karakteristik yang sama atau mirip dengan kelompok yang lain tetapi dengan kelompok yang lain memiliki karakteristik yang berbeda [9]

2.5. Minat

Minat adalah perasaan tertarik terhadap suatu hal atau kegiatan tertentu. Peduli pada dasarnya berarti menerima sesuatu antara diri sendiri dan orang luar, semakin kuat atau dekat hubungannya, semakin besar minatnya. Oleh karena itu, minat dikatakan bahwa preferensi adalah keinginan untuk menjalin hubungan dengan lingkungan, kecenderungan untuk memeriksa, menyelidiki, atau melakukan aktivitas yang diminati seseorang [10].

3. METODE PENELITIAN



Gambar 1. Alur penelitian

Metode yang digunakan dalam penelitian ini menggunakan metode eksperimen yaitu metode yang bertujuan untuk menguji pengaruh variabel yang satu terhadap variabel yang lain atau untuk menguji apakah hubungan sebab akibat antara variabel yang satu dengan variabel yang lain [11]. Pada penelitian ini metode eksperimen digunakan untuk mengelompokkan minat siswa dan mengetahui parameter mana yang paling baik digunakan oleh algoritma K-Means.

3.1. Pengumpulan Data

a. Wawancara

Dalam penelitian ini pada wawancara dilakukan bersama dengan salah satu staff kesiswaan dan juga staff bursa kerja khusus (BKK). Wawancara ini dilakukan di SMK PUI Gegesik

b. Angket

cara ini dilakukan dengan cara mengajukan daftar pertanyaan yang harus di jawab oleh responden yaitu peserta didik SMK PUI Gegesik

c. Studi Dokumentasi

Pada tahapan ini memperoleh data dan informasi berupa catatan tertulis atau gambar yang berkaitan dengan masalah ini, data tersebut berupa data yang berasal dari jurnal-jurnal yang relevan, buku, serta informasi hasil wawancara terhadap narasumber

3.2. Data Selection

Dalam penelitian ini proses data selection akan melakukan seleksi pada data set dengan menggunakan correlation matrix untuk mengetahui atribut yang bisa digunakan dalam proses berikutnya.

3.3. Data Preprocessing

Pada proses preprocessing data, data akan disiapkan untuk diproses jadi dilakukan mengatur role atribut seperti halnya atribut NISN akan di jadikan ID untuk menggantikan atribut nama, serta pemilihan

atribut menggunakan operator select attributes yang dimana atribut akan dipilih.

3.4. Data Transformation

Dalam proses Transformation data, data yg non-numerik seperti Jurusan, Jenis Kelamin, Minat, Tanggal lahir, Personaliti akan di ubah menjadi numerik dengan menggunakan operator *nominal to numerical*.

3.5. Data Mining

Pada pengelompokan ini parameter yang dipilih ada 3 yaitu *Clustering K* untuk mengetahui nilai K, *clustering max runs* untuk mengatur iterasi yang dipakai dan *clustering measure type* untuk memilih *measure* apa saja yang akan digunakan, dalam hal ini peneliti menggunakan 3 *measure type* yaitu *Mixed Measure*, *Numerical Measure* dan *Bregman Divergences*

3.6. Evaluation

Pada tahap ini data yang selesai diolah akan ditampilkan hasilnya, dalam penelitian ini tahap evaluation digunakan sebagai pemaparan hasil dari 3 *measure* yang digunakan berdasarkan dari nilai K, iterasi, dan DBI

3.7. Kesimpulan

Pada tahap akhir hasil penelitian dan pengujian merupakan proses pengkajian ulang terhadap hasil penelitian dan bisa di tarik kesimpulan dari penelitian

4. HASIL DAN PEMBAHASAN

4.1. Data Set

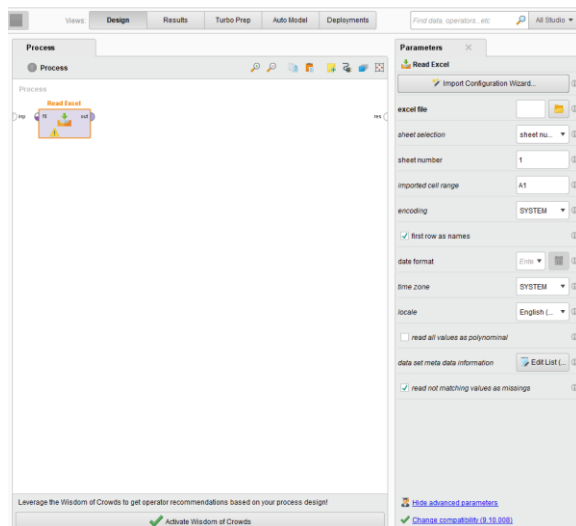
Data set yg di pada penelitian ini di dapat dari SMK PUI Gegesik melalui wawancara secara langsung dengan pihak BKK dan Kesiswaan, data yang di peroleh dari SMK PUI Gegesik data tersebut berupa nilai dan minat siswa. Data yang di dapat sebanyak 795 data siswa SMK PUI Gegesik.

Table 1. sample data set

No	nama	Jenis kelamin	NISN	Kelas	Jurusan	Tempat Lahir	Tanggal Lahir	personalit y	Nilai Akademik	minat
1	Asriatul dairoh	P	0065772319	10	Multimedia	Cirebon	2006-10-20	Baik	84	Kerja
2	Dewi Ramadhani	P	0059382942	10	Multimedia	Cirebon	2005-10-24	Baik	80	Kerja
3	Fadhilah Nur ahyari	P	0066518737	10	Multimedia	Cirebon	2006-08-04	Baik	85	Kerja
794	Wahyu Ali Sigit	L	0037602580	12	Teknik Pengelasan	Cirebon	2003-11-03	Cukup	78	Belum Menentukan
795	Yanwar Hadi	L	43053886	12	Teknik Pengelasan	Cirebon	2004-01-08	Cukup	70	Kerja

Dari data yang di dapat penulis, data tersebut mempunyai 10 atribut diantaranya yaitu nama, jenis kelamin, NISN, kelas, jurusan, tempat lahir, tanggal lahir, personality, nilai akademik, dan minat.

Pada langka pertama yang dilakukan adalah *importing data* maka diperlukan sebuah operator. Berhubung data yang digunakan berformat excel (xlsx), maka operator yang dibutuhkan adalah *Read Excel*.



Gamba 2. Read excel

Jika sudah maka klik pada *Import Configuration Wizard* seperti di Gambar 2, lalu cari *file data set* yang berformat xlsx yang akan diproses.

Selanjutnya pilih pada *table* atribut apa saja yang akan diproses lalu pilih next, Jika data set tidak terjadi masalah maka pilih finish.

4.2. Data Selection

Pada proses ini data akan diseleksi berdasarkan atribut di dalam dataset menggunakan operator correlation matrix untuk mengetahui atribut yang bisa digunakan, data yg akan diproses adalah data rekap nilai dan minat siswa.

Salah satu cara untuk menghitung hubungan ini adalah dengan menggunakan koefisien korelasi Pearson, yang merupakan ukuran hubungan linier antara dua variabel. Ini memiliki nilai antara -1 dan 1 di mana:

- -1 menunjukkan korelasi linier negatif sempurna antara dua variable
- 0 menunjukkan tidak ada korelasi linier antara dua variable
- 1 menunjukkan korelasi linier positif sempurna antara dua variable

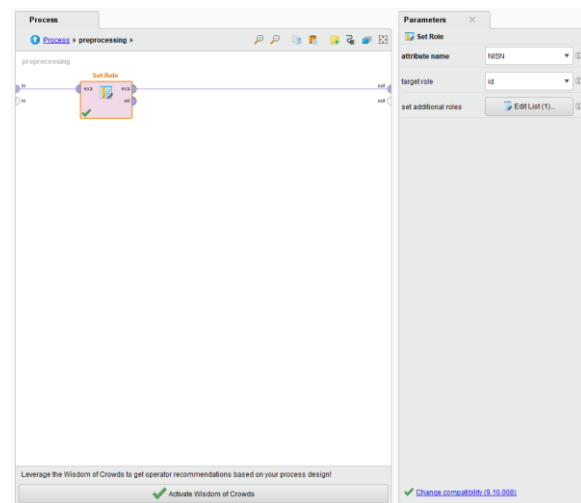
Attribut...	Nama	L/P	Jurusan	Tempat...	Tanggal...	Person...	NISN	Kelas	Nilai Ak...
Nama	1	0.120	0.529	-0.026	0.877	0.094	0.008	0.940	-0.009
L/P	0.120	1	0.366	0.014	0.103	0.336	0.049	0.060	-0.267
Jurusan	0.529	0.366	1	-0.019	0.441	0.199	0.042	0.367	-0.121
Tempat...	-0.026	0.014	-0.019	1	-0.012	-0.007	0.024	-0.039	-0.036
Tanggal...	0.877	0.103	0.441	-0.012	1	0.093	-0.001	0.840	-0.001
Personali...	0.094	0.336	0.199	-0.007	0.093	1	-0.009	0.088	-0.206
NISN	0.008	0.049	0.042	0.024	-0.001	-0.009	1	-0.021	-0.043
Kelas	0.940	0.060	0.367	-0.039	0.840	0.088	-0.021	1	0.033
Nilai Aka...	-0.009	-0.267	-0.121	-0.036	-0.001	-0.206	-0.043	0.033	1

Gambar 3. Hasil Correlation Matriks

Semakin jauh koefisien korelasi dari nol, semakin kuat hubungan antara kedua variabel tersebut. Hasil dari correlation matrix terlihat atribut yang bisa digunakan yaitu nama, jenis kelamin, jurusan, tanggal lahir, personality, Nisn, kelas. Dan untuk atribut yang tidak bisa di pakai yaitu atribut nilai akademik dan tempat lahir, karena terlalu banyak nilai (-) yang berkaitan dengan atribut lain yang berarti atribut tersebut tidak bisa di pakai

4.3. Data Preprocessing

Proses selanjutnya setelah melakukan seleksi data maka data akan dipersiapkan dengan mengatur *role* dalam data. Untuk mengatur *role* data dalam rapidminer studio perlu menggunakan operator *set role*, dalam tahap ini *role* NISN akan dijadikan ID menggunakan operator *set role*.



Gambar 4. Set Role

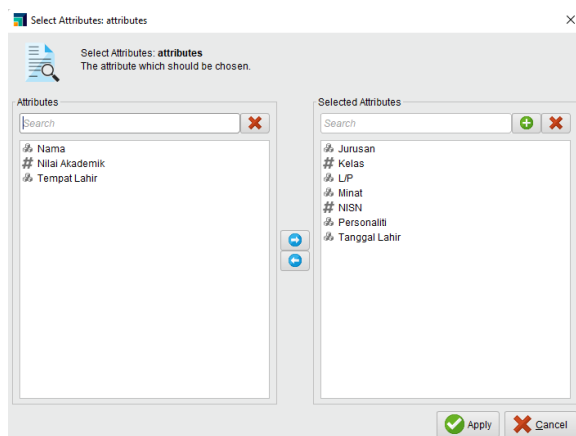
Pada *attributes name* pilih atribut yang akan diubah *rolenya* lalu pilih pada target *rolenya* dengan *role* yang akan ditetapkan pada atribut yang dipilih, pada pengelompokan ini atribut NISN akan dijadikan ID untuk mengganti atribut nama.

Name	Type	Missing	Stat...	Filter (10 / 10 attributes)	Search for Attributes
NISN	Real	0	Min: 9860803, Max: 3137664162, Average: 10737...		
Nama	Polynomial	0	Least: sunandi (1), Most: ADITYA (2), Values: ADITY...		
LIP	Polynomial	0	Least: P (301), Most: L (494), Values: L (494...		
Kelas	Integer	0	Min: 10, Max: 12, Average: 10.940		
Jurusan	Polynomial	0	Least: Teknik Logistik (92), Most: Multimedia (214), Values: Multime...		
Tempat Lahir	Polynomial	0	Least: TEGAL (1), Most: Cirebon (755), Values: Cirebon...		
Tanggal Lahir	Polynomial	0	Least: 38993 (1), Most: 2005-12-25 (5), Values: 2005-1...		
Personaliti	Polynomial	0	Least: Cukup (229), Most: Baik (566), Values: Baik (5...		
Nilai Akademik	Integer	0	Min: 70, Max: 93, Average: 78.214		
Minat	Polynomial	0	Least: Wirausaha (62), Most: Kerja (495), Values: Kerja (...		

Showing attributes 1 - 10 Examples: 795 Special Attributes: 1 Regular Attributes: 9

Gambar 5. Hasil Set Role

Terlihat hasil *set role* pada bagian kanan bawah terdapat 9 atribut regular dan 1 atribut *special*. Jika sudah maka proses selanjutnya dalam *preprocessing* adalah melakukan pemilihan atribut, pada tahap ini dibutuhkan operator dalam rapidminer studio yaitu operator *select attributes*.



Gambar 6. Select Attributes

Jika sudah ditambahkan maka kita perlu mengganti pada bagian *attributes filter type* menjadi subset agar bisa memilih atribut apa saja yang akan kita pilih. Pemilihan atribut diatas berdasarkan hasil uji *correlation matrix* pada tahap *data selection* dan atribut nama juga digantikan dengan NISN jadi atribut nama tidak diperlukan.

Name	Type	Missing	Stat...	Filter (7 / 7 attributes)	Search for Attributes
NISN	Real	0	Min: 9860803, Max: 3137664162, Average: 10737...		
LIP	Polynomial	0	Least: P (301), Most: L (494), Values: L (494...		
Kelas	Integer	0	Min: 10, Max: 12, Average: 10.940		
Jurusan	Polynomial	0	Least: Teknik Logistik (92), Most: Multimedia (214), Values: Multime...		
Tanggal Lahir	Polynomial	0	Least: 38993 (1), Most: 2005-12-25 (5), Values: 2005-1...		
Personaliti	Polynomial	0	Least: Cukup (229), Most: Baik (566), Values: Baik (5...		
Minat	Polynomial	0	Least: Wirausaha (62), Most: Kerja (495), Values: Kerja (...		

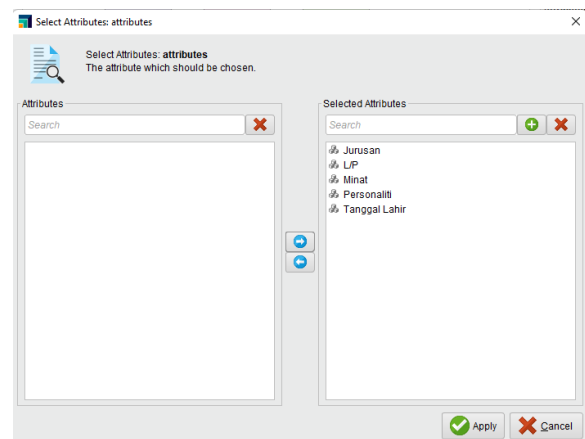
Showing attributes 1 - 7 Examples: 795 Special Attributes: 1 Regular Attributes: 6

Gambar 7. Hasil Select Attributes

Pada bagian ini atribut yang dipilih adalah 6 dan 1 atribut spesial yaitu NISN dan atribut regular seperti Jurusan, Kelas, Jenis Kelamin, Minat, Personaliti dan tanggal lahir.

4.4. Data Transformation

Jika sudah selesai pada tahapan *preprocessing* data maka selanjutnya melakukan *transformation data*, karena pada proses *data mining* hanya bisa membaca *type data numerik* maka diperlukan operator untuk mengubah *type data non-numerik* menjadi numerik yaitu operator *nominal to numerical* [12].



Gambar 8. Transformation Data

Sama halnya dengan *select attributes* pada bagian *attributes filter type* diubah menjadi *subset* agar bisa memilih atribut yang akan dipilih, pada tahap ini *type data* yang non-numerik akan tersaring dan perlu diubah menjadi *type data numerik*, terdapat 6 atribut yang perlu diubah ke numerik yaitu Jurusan, Jenis Kelamin, Minat, Personaliti dan tanggal lahir.

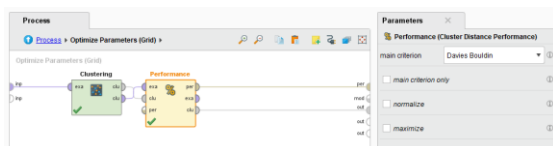
Row No.	NISN	L/P	Jurusan	Tanggal Lahir	Personaliti	Minat	Kelas
1	65772319	0	0	0	0	0	10
2	56362942	0	0	1	0	0	10
3	66518737	0	0	2	0	0	10
4	64403790	0	0	3	0	0	10
5	52468150	1	0	4	0	1	10
6	55486424	1	0	5	1	0	10
7	63180348	0	0	6	0	0	10
8	62787448	0	0	7	0	0	10
9	55524889	0	0	8	0	0	10
10	62871984	0	0	9	0	0	10
11	58135615	0	0	10	0	2	10
12	59508710	1	0	11	0	0	10
13	65428681	1	0	12	1	0	10
14	66643958	1	0	13	0	0	10
15	51837668	1	0	14	1	0	10
16	68777178	1	0	15	0	0	10
17	67783048	0	0	13	0	0	10
18	64715548	0	0	16	0	1	10
19	48113855	0	0	17	0	2	10
20	58407815	0	0	18	0	0	10
21	3065733971	1	0	19	0	0	10
22	63489857	1	0	20	0	0	10
23	64697048	1	0	21	0	0	10

Gambar 9. Hasil Transformation Data

Seperti yang terlihat pada Gambar 10 itu adalah hasil dari operator *nominal to numerical* data yang tadi bertipe data non-numerik berubah menjadi numerik. Dan jika sudah bertipe data numerik maka *dataset* sudah siap untuk proses *data mining*.

4.5. Data Mining

Pada proses data mining ini akan menggunakan operator tambahan seperti *Optimize parameter* (Grid), karena untuk mengetahui nilai optimal dari suatu parameter dalam pengelompokan ini. Didalam *Optimize parameter* pada subprosesnya perlu ditambahkan algoritmanya



Gambar 10. Optimize parameter

Disini menggunakan algoritma *K-means* dan untuk segmentasi menggunakan *cluster distance performance*. Lalu pada bagian *main criterion* menggunakan *Davies bouldin* karena untuk mengetahui nilai Dbinya.

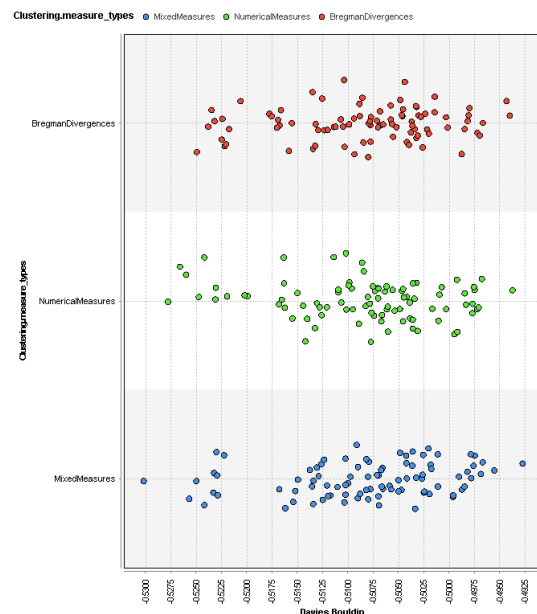
4.6. Evaluation

Table 2. Hasil 3 measure

K	Mixed Measure	Numerical Measure	Bregman Disvergences
K=2	-0.497	-0.497	-0.497
K=3	-0.502	-0.502	-0.502
K=4	-0.522	-0.522	-0.522
K=5	-0.509	-0.508	-0.508
K=6	-0.497	-0.504	-0.503
K=7	-0.508	-0.509	-0.508
K=8	-0.510	-0.510	-0.513
K=9	-0.503	-0.504	-0.503
K=10	-0.499	-0.499	-0.502

Pada table diatas terlihat perbandingan hasil dari K=2 sampai K=10 dengan nilai terbaik yang sama yaitu 0,497 yang dimana perbandingan dari 3 measure tersebut relative tipis, jadi diperlukan scater untuk menampilkan perbedaan jarak antara 3 measure tersebut

4.7. Pembahasan

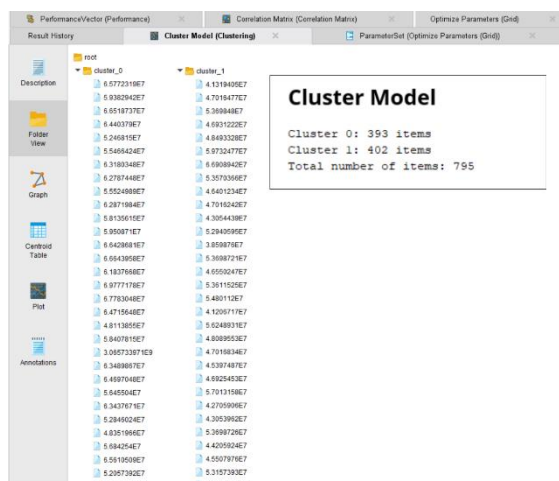


Gambar 11. Perbandingan scater 3 measure

Pada Gambar 11 Terlihat hasil nilai DBI dari setiap measure berbeda dan nilai terbaik terletak pada Mixed Measure sedangkan Numerical measure tepat bedekatan dengan nilai DBI dari Mixed measure lalu terakhir adalah bregman divergences.

DBI adalah alat untuk mengukur validitas cluster dalam metode clustering. Semakin kecil nilai DBI (DBI mendekati 0) maka hasil cluster akan semakin optimal.[4] Untuk k = 2, jenis measure = mixed measure memiliki nilai DBI terdekat dengan 0, yaitu 0,497. diketahui bahwa parameter yang terbaik dalam pengelompokan data minat siswa di SMK PUI Gegesik adalah menggunakan Mixed Measure sebagai parameter untuk mendapatkan nilai optimal.

Hasil kluster dari model proses ini mendapatkan 2 kluster set yang dimana 2 kluster set itu hasil default kluster dari rapid miner sendiri jika tidak mengatur berapa jumlah kluster set yang kita inginkan, dan disini terdiri dari 2 kluster set yaitu cluster 0 yang mempunyai jumlah items 393 dan cluster 1 dengan jumlah items 402. Cluster 0 sendiri merujuk ke nilai tertinggi dan cluster 1 merujuk ke nilai terendah dan jumlah total item dari ke 2 cluster model tersebut yaitu 795 tersebut yaitu 795.



Gambar 12. hasil cluster

5. KESIMPULAN DAN SARAN

Dalam penelitian ini, kami mengusulkan metode k-means yang dioptimalkan oleh DBI untuk menentukan jumlah cluster yang optimal. Dengan melakukan uji beda dengan jumlah cluster ($k = 2, 3, 4, 5, 6, 7, 8, 9, 10$), hasil metode k-means dengan tipe measure adalah Mixed Measures dengan DBI paling optimal sebesar 0,497 dengan 2 cluster.

DAFTAR PUSTAKA

- [1] S. Aulia, "KLASTERISASI POLA PENJUALAN PESTISIDA MENGGUNAKAN METODE K-MEANS CLUSTERING (STUDI KASUS DI TOKO JUANDA TANI KECAMATAN HUTABAYU RAJA)," *Djtechno J. Teknol. Inf.*, vol. 1, no. 1, 2021, doi: 10.46576/djtechno.v1i1.964.
- [2] F. Firza and S. Sarjono, "Penerapan Algoritma K-Means Dalam Metode Clustering Untuk Peminatan Jurusan Bagi Siswa Swasta Pelita Raya Kota Jambi," *J. Manaj. Sist. Inf.*, vol. 5, no. 3, pp. 371–382, 2020, [Online]. Available: <http://ejournal.stikom-db.ac.id/index.php/manajemensisteminformasi/article/view/907>
- [3] M. Arista Ulfa and M. Hidayat, "Product Clustering Analysis ... (Maria Arista Ulfa, et al.) PRODUCT CLUSTERING ANALYSIS ON THE MARKETPLACE USING K-MEANS APPROACH (CASE STUDY: SHOPEE)." [Online]. Available: <http://journal.ugm.ac.id/index.php/ajse>
- [4] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM J.*, vol. 10, no. 3, pp. 1099–1103, 2021, doi: 10.18421/TEM103-13.
- [5] S. Butsianto and N. Saepudin, "Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 1, 2020, doi: 10.32672/jnkti.v3i1.2008.
- [6] T. Purwanti, W. Ramdhan, and S. Santoso, "Penerapan Metode Klasterisasi K-Means untuk Strategi Promosi Pada SMK Tamansiswa Sukadama," *JUTSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 2, no. 1, 2021, doi: 10.33330/jutsi.v2i1.1156.
- [7] R. Muliono and Z. Sembiring, "Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 4, no. 2, 2019.
- [8] W. A. Suputra, "Klasterisasi Hasil Ujian Nasional SMA/MA dengan Algoritma K-Means," *Wahana Mat. dan Sains J. ...*, vol. 15, no. 1, 2021.
- [9] Y. A. Wijaya et al., "K-Means Di Sekolah Menengah Kejuruan Wahidin Kota Cirebon," vol. 6, no. 2, pp. 552–559, 2022.
- [10] G. Mussardo, "Pengaruh Pemanfaatan Teori Holland Dalam Layanan Informasi Terhadap Minat Karir Siswa Di Sekolah Menengah Atas Negeri 3 Singingi Hilir," *Stat. F. Theor.*, vol. 53, no. 9, pp. 1689–1699, 2019.
- [11] Ahmad Fauzi, U. Hayati, and F. M. Basysyar, "Perancangan Aplikasi Point of Sales Menggunakan Android Native Di Ud. Murti Aji Cirebon," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 544–551, 2022, doi: 10.36040/jati.v6i2.5250.
- [12] M. Ripai, U. Hayati, W. Widyawati, and H. Susana, "KOPERTIP: Jurnal Ilmiah Manajemen Informatika dan Komputer Pengklasifikasian Surat Pemberitahuan Pajak Daerah Menggunakan Metode Regresi Logistik Biner Untuk Mengetahui Patuh Dan Tidak Patuh Dalam Pembayaran Pajak Daerah," vol. 06, no. 01, pp. 27–33, 2022, [Online]. Available: <http://jurnal.kopertipindonesia.or.id/>