Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 6 ; Nomor 1 ; Januari 2023 ; Page 235-245
E-ISSN : 2615-5133 ; **P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

Implementasi Data Mining Untuk Rekomendasi Jurusan Menggunakan Algoritma K-Means Clustering

Muhammad Syahril[1], Sri Kusnasari[2], Sobirin[3], Abdullah Muhazir[4], Astri Syahputri[5]

[1,2,3,4,5] Sistem Informasi, Stmik Triguna Dharma
Email: [1]muhammadsyahril.tgd@gmail.com, [2]srikusnasari.tgd@gmail.com, [3]Sobirin1104@yahoo.co.id, [4]muhazir@gmail.com, [5]astri.syahputribakpaw@gmail.com
Email Penulis Korespondensi: muhammadsyahril.tgd@gmail.com

Article History:
Received Dec 12th, 2022
Revised Jan 12th, 2023
Accepted Jan 21th, 2023

Abstrak

Penjurusan merupakan hal yang wajib dan tidak dapat terelakkan dalam dunia pendidikan termasuk pada saat sekolah menengah atas (SMA).Pada SMA Negeri 2 Kabanjahe pemilihan jurusan dilakukan pada saat berada di kelas X (sepuluh). Hak dalam pemilihan jurusan itu sendiri diberikan langsung kepada siswa. Hal ini membuat banyak siswa mengalami kebingungan dalam menentukan jurusan yang akan dipilih karena adanya beberapa faktor yang mempengaruhi, diantaranya kurangnya pemahanan siswa terhadap jurusan tersebut, tuntutan orangtua bahkan sekedar mengikuti teman. Nilai akademis dapat diajukan sebagai acuan untuk mengetahui jurusan apa yang tepat bagi siswa berdasarkan nilai akademis. Hal ini membuat peneliti melakukan penelitian pada SMA Negeri 2 Kabanjahe terkait bagaimana memberikan rekomendasi penjurusan. Salah satu tehnik dalam data mining yang dapat dilakukan untuk pengelompokan yaitu clustering. Hasil penelitian ini berupa sistem yang dapat memberikan rekomendasi jurusan pada siswa baru menggunakan K-Means.

Kata Kunci : Penjurusan, Siswa, Sekolah, Data Mining, Clustering

Abstract

Majors are mandatory and unavoidable in the world of education, including during high school (SMA). At SMA Negeri 2 Kabanjahe the choice of majors is made while in class X (ten). The right to choose the major itself is given directly to students. This makes many students experience confusion in determining which major to choose because of several influencing factors, including students' lack of understanding of the major, parents' demands and even just following friends. Academic scores can be submitted as a reference to find out what majors are right for students based on academic grades. This prompted researchers to conduct research at SMA Negeri 2 Kabanjahe regarding how to provide major recommendations. One technique in data mining that can be done for grouping is clustering. The results of this study are in the form of a system that can provide major recommendations to new students using K-Means.

Keyword : *Majors, Students, Schools, Data Mining, Clustering*

1. PENDAHULUAN

Pendidikan merupakan salah satu faktor yang penting dalam menentukan kemajuan suatu bangsa, karena dengan adanya pendidikan dapat mengembangkan berbagai potensi yang ada dalam diri seseorang. Proses pendidikan di sekolah menengah atas merupakan suatu proses yang memiliki pengaruh yang sangat besar terhadap perkembangan siswa di masa depan. Salah satu proses yang dilaksanakan pada sekolah menengah atas (SMA) adalah penjurusan siswa [1]. Pengambilan keputusan dalam pemilihan jurusan diberikan langsung kepada siswa, dimana penjurusan pada siswa SMA dilakukan pada saat siswa berada di kelas sepuluh [2].

Dalam kurikulum penjurusan di SMA hanya memiliki 3 jurusan diantaranya IPA, IPS, dan Bahasa. SMA Negeri 2 Kabanjahe merupakan salah satu sekolah yang hanya memiliki 2 jurusan untuk dipilih yaitu IPA dan IPS. Jurusan IPA dan IPS ini merupakan jurusan yang paling umum di sekolah menengah atas (SMA). Mata pelajaran jurusan IPA lebih

menitikberatkan pada penguasaan konsep-konsep IPA untuk kepentingan siswa menyelesaikan masalah dalam kehidupan sehari-hari. Sedangkan untuk jurusan IPS lebih menitikberatkan pengembangan keterampilan ilmu sosial.

Penjurusan sangat penting bagi siswa selain agar siswa dapat menikmati proses belajar mengajar juga sangat penting karena dapat menjadi suatu acuan bagi masing-masing siswa untuk melanjutkan studi kasus di masa depan. Siswa SMA memang cenderung akan melajutkan studi ke perguruan tinggi setelah lulus SMA, berbeda dengan siswa SMK dimana mereka mengambil kejuruan untuk mendapatkan kemampuan tertentu dalam satu bidang untuk langsung bekerja dengan kemampuan yang mereka punya. Namun bukan berarti tidak ada siswa SMK yang tidak melanjutkan studi ke perguruan tinggi. Jurusan pada saat SMA berkaitan dengan jurusan yang akan dilanjutkan pada perguruan tinggi. Maka dari itu menentukan jurusan pada saat SMA merupakan hal yang penting untuk kenyamanan siswa dalam proses belajar dan juga untuk studi selanjutnya.

Dalam menentukan jurusan para siswa juga memiliki kesulitan dalam menentukan jurusan mana yang tepat bagi mereka. Beberapa faktor yang memengaruhi siswa yaitu kurangnya pemahaman akan jurusan tersebut, tuntutan orang tua, bahkan sekedar mengikuti teman. Kesalah siswa dalam memilih jurusan memiliki dampak yang cukup besar, siswa akan mudah stress karena kurang memahami pelajaran pada jurusan tersebut sehingga mengganggu proses belajar mengajar. Kesalahan ini bahkan dapat menggangu studi lanjut siswa dimasa depan karena individu cenderung melanjutkan pendidikan hingga perguruan tinggi sehingga memudahkan mendapatkan pekerjaan yang layak demi masa depannya. Penentuan jurusan saat ini memiliki banyak kelemahan, antara lain hanya berdasarkan keinginan siswa tanpa melihat latar belakang siswa, seperti nilai akademisnya. Nilai akademis dapat dijadikan sebagai acuan untuk mengetahui jurusan mana yang tepat bagi siswa. Dalam hal ini diperlukan sistem yang dapat memberikan rekomendasi jurusan kepada siswa berdasarkan nilai akademis. Berangkat dari permasalahan di atas, penelitian ini akan menggunakan data siswa kelas sepuluh SMA Negeri 2 Kabanjahe. Dengan menggunakan metode *K-Means* dalam peng-*claster*-an yang akan menghasilkan informasi rekomendasi jurusan siswa baru pada SMA Negeri 2 kabanjahe. Selanjutnya algoritma yang akan digunakan untuk pengolaan *Data Mining* pada kasus menganalisa data siswa dan digunakan untuk pembentukan *cluster* adalah algoritma *K-Means*.

Data mining adalah proses yang menggunakan tehnik *statistic*, matematika, kecerdasan buatan, *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari *database* yang besar [3]. Secara teknis, *data mining* dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan *field* dari sebuah rasional database yang besar. Penerapan *K-means* sudah dilakukan oleh peneliti sebelumnya, diantaranya untuk Pengelompokan Kasus dan Kematian COVID-19 di Asia Tenggara menggunakan Algoritma K-Means [4], untuk Pengelompokan Rekomendasi Tugas Akhir [5] dan untuk Mengelompokkan Penjualan Produk [6]. *K-Means Clustering* merupakan suatu algoritma pengelompokan yang cukup sederhana yang mempartisi *database* kedalam beberapa *cluster* k [7].

Penerapan ilmu *data mining* dalam pengelompokan data dengan menggunakan metode *K-Means* akan bertujuan untuk mempermudah siswa dalam mengambil keputusan berdasarkan rekomendasi jurusan dengan cepat dan tepat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Metode penelitian memberikan gambaran rancangan penelitian yang meliputi antara lain:

a. Tehnik Pengumpulan Data

Pengumpulan data adalah tehnik atau cara yang dilakukan oleh peneliti untuk mengumpulkan data. Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian.

1. Wawancara dengan pihak sekolah
2. Observasi langsung ke SMA Negeri 2 Kabanjahe

b. Studi Literatur

Studi Kepustakaan merupakan salah satu elemen yang mendukung sebagai landasan teoritis penelitian untuk mengkaji masalah yang dibahas, melalui beberapa jurnal dan refrensi lainnya.

2.2 Data *Mining*

Data mining merupakan proses untuk menggali nilai tambah dari sekumpulan data berupa pengetahuan yang selama ini tersembunyi atau tidak diketahui secara manual [8]. *Data mining* disebut juga dengan *knowledge discovery in database* (KKD) ataupun *pattern recognition*. KDD *data mining* digambarkan sebagai proses pencarian pengetahuan yang menarik dalam *database* seperti pola, asosiasi, aturan, perubahan, keganjilan dan struktur penting dari sejumlah besar bank data dan tempat penyimpanan informasi lainnya [9].

2.3 Algoritma K-Means

Algoritma *K-Means* merupakan algoritma pengelompokan *iterative* yang melakukan partisi set data ke dalam sejumlah K *cluster* yang sudah ditetapkan di awal [10]. Algoritma *K-Means* sederhana untuk diimplementasikan dan dijalankan, *relative* cepat, mudah beradaptasi, umum penggunaaannya dalam praktek. Bahkan algoritma *clustering*

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 6 ; Nomor 1 ; Januari 2023 ; Page 235-245
E-ISSN : 2615-5133 ; **P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

berusaha mengsegmentasi seluruh kumpulan data ke dalam subkelompok-subkelompok atau *cluster-cluster homogeny* secara *relative* [11].

Algoritma *K-Means Clustering* adalah algoritma sederhana dan efektif untuk menemukan *cluster* dalam data dengan algoritma sebagai berikut : [12]

1. Tentukan jumlah *cluster*
2. Tentukan nilai yang menjadi lokasi pusat *cluster* awal
3. Hitung pusat *cluster* terdekat untuk setiap record
4. Untuk setiap *cluster* k, hitung *centroid cluster* dan memperbaharui lokasi setiap pusat *cluster*
5. Ulangi langkah 3 sampai 5 sampai konvergengsi atau penghentian

3. HASIL DAN PEMBAHASAN

3.1 Penerapan K-Means

Berikut langkah-langkah pada algoritma *K-Means* sampai diketahui pembagian nilai *centroid* sebelumnya tidak berubah, sesuai pada tabel 1.

1. Tentukan jumlah *cluster*, *cluster* k = 2
2. Tentukan nilai yang menjadi lokasi titik awal *cluster*

Tabel 1. Data *Centroid* Awal

Titik Pusat *Centroid*	Nilai Awal *Cluster*					
Data ke-9 (C1)	93	93	90	89	90	87
Data ke-3 (C2)	68	69	68	75	80	78

3. Hitung pusat *cluster* terdekat untuk setiap *record*

a. Perhitungan jarak dari data ke-1 dengan pusat *cluster*

$$D(1,1) = \sqrt{(90-93)^2} + (87-93)^2 + (88-90)^2 + (80-89)^2 + (84-90)^2 + (80-87)^2$$
$$D(1,1) = 33$$

$$D(1,2) = \sqrt{(90-68)^2} + (87-69)^2 + (88-68)^2 + (80-75)^2 + (84-80)^2 + (80-78)^2$$
$$D(1,2) = 71$$

b. Perhitungan jarak dari data ke-2 dengan pusat *cluster*

$$D(2,1) = \sqrt{(95-93)^2} + (90-93)^2 + (90-90)^2 + (81-89)^2 + (79-90)^2 + (69-87)^2$$
$$D(2,1) = 42$$

$$D(2,2) = \sqrt{(95-68)^2} + (90-69)^2 + (90-68)^2 + (81-75)^2 + (79-80)^2 + (69-78)^2$$
$$D(2,2) = 86$$

c. Perhitungan jarak dari data ke-2 dengan pusat *cluster*

$$D(2,1) = \sqrt{(68-93)^2} + (69-93)^2 + (68-90)^2 + (75-89)^2 + (80-90)^2 + (78-87)^2$$
$$D(2,1) = 104$$

$$D(2,2) = \sqrt{(68-68)^2} + (69-69)^2 + (68-68)^2 + (75-75)^2 + (80-80)^2 + (78-78)^2$$
$$D(2,2) = 0$$

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 6 ; Nomor 1 ; Januari 2023 ; Page 235-245
E-ISSN : 2615-5133 ; **P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

Untuk lebih lengkapnya jarak pada setiap baris data, hasilmya seperti pada tabel 2 berikut:

Tabel 2. Hasil Perhitungan Iterasi 1

Nama	Nilai Jarak		Jarak Terdekat	Jarak Terdekat^2
	C1	C2		
Ade Depari	33	71	33	1089
Aldi Ahmad Dhani	42	86	42	1764
Aldo Girsang	104	0	0	0
Alfi Syahra Ginting	43	61	43	1849
Amista Ginting	107	7	7	49
Anas Tasya	93	29	29	841
Angel Tridhany	41	63	41	1681
Anggi Ardian	71	45	45	2025
Anisa Ginting	0	104	0	0
Ardiani Sembiring	39	67	39	1521
Arjuna Ginting	50	54	50	2500
Aulia Purba	32	74	32	1024
Aurelia Cinta Uli	75	29	29	841
Bastanta Silangit	36	70	36	1296
Brando Sirait	41	77	41	1681
Brian Depari	91	17	17	289
Bripora Tarigan	61	43	43	1849
Caca Yustika	39	77	39	1521
Cesylia Tamba	41	63	41	1681
Chika Trivena Marbun	40	64	40	1600
ClaraBenedita Sinuhaji	71	35	35	1225
Claudia Silitonga	66	38	38	1444
Crisda Stephani	116	26	26	676
Dandi Lucky Cristoper	74	36	36	1296
Dear Meliani Tarigan	64	40	40	1600
Debora Tania	120	20	20	400
Diego Karo	14	90	14	196
Dira Aulia	123	19	19	361
Divia Tesalonika T	46	58	46	2116
Edi Suranta Tarigan	69	35	35	1225
Egi Depari	49	55	49	2401
Elfransius Sembiring	107	15	15	225
Elpa Tarigan Girsang	23	95	23	529
Emelia Teodora	92	12	12	144
Emiya Pepayosa Giting	147	43	43	1849
Eriska Perangin Angin	65	39	39	1521
Etisa Sinulaki	71	55	55	3025
Evrani Tabina	60	44	44	1936
Febry Yhana Laia	16	94	16	256
Febyola Tarigan	71	55	55	3025
Finka Sihombing	108	30	30	900
Friska Natalia	39	69	39	1521
Gebita Putri Melinia	75	47	47	2209
Geby Marselina	91	31	31	961
Gia Kinarian	11	109	11	121
Gilang Saragih	38	66	38	1444
Gobin Ginting	66	46	46	2116
Grace Marpaung	55	51	51	2601
Hafifah Cairyah	113	13	13	169

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 6 ; Nomor 1 ; Januari 2023 ; Page 235-245
E-ISSN : 2615-5133 ; **P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

Herdianta Tarigan	120	30	30	900
Pebiola Br Sembiring	47	57	47	2209
Petrus Siagian	61	45	45	2025
Piere Traigan	12	96	12	144
Priadi Sihombing	54	50	50	2500
Prilly Dwi Gita	31	77	31	961
Purnama Sari	14	98	14	196
Rama Dania	42	70	42	1764
Rani Lestari Sitepu	57	49	49	2401
… dst				
Zidane Sion Ruben	71	37	37	1369
	WCV			**30.426.487**

Dari tabel 2 didapat hasil *cluster* sebagai berikut:

C1:
{1,2,4,7,9,10,11,12,13,14,15,18,19,20,27,29,31,33,39,42,45,46,52,56,58,60,61,65,
70,71,72,77,83,85,86,87,91,94,98,102,105,107,109,110,111,114,117,120,124,125,127,131,141,143,145,147}
C2:
{3,5,6,8,13,16,17,21,22,23,24,25,26,28,30,32,34,35,36,37,38,40,41,43,44,47,48,49,50,51,53,54,55,57,59,62,63,64,66,67,68,69,73,74,75,76,78,79,80,81,84,88,89,90,92,93,95,97,99,100,101,103,104,106,108,112,113,115,116,118,119,121,122,123,126,129,130,132,133,134,135,136,137,138,139,140,142,144,146,148,149,150}

4. Menghitung Nilai BCV (*Between Cluster Variation*)
Karena *cemtroid* m1 (93,93,90,89,90,87), m2(68,69,68,75,80,78)

D(m1,m2) $= \sqrt{(93-68)^2} + (93-69)^2 + (90-68)^2 + (89-75)^2 + (90-80)^2 + (87-78)^2$
$= 104$
BCV $= 104$

5. Menghitung Besar Rasio
Rasio besaran antara BCV *(Between Cluster Variation)* dengan WCV (*Within Cluster Variation*)

Besar Rasio = BCV/WCV
= 104/197157
= 0,0005274984

Karena ini merupakan iterasi ke-1, maka dilanjutkan ke Iterasi selanjutnya. Pembaruan *centroid* dengan menghitung rata-rata nilai pada masing-masing *cluster* tertera pada tabel 3.

Tabel 3. Hasil Perhitungan Iterasi 2

Nama	Nilai Jarak		Jarak Terdekat	Jarak Terdekat^2
	C1	C2		
Ade Depari	20,43638	46,37894	20,43638	417,6456
Aldi Ahmad Dhani	45,43638	61,27368	45,43638	2064,4646
Aldo Girsang	77,21818	26,7263	26,7263	714,29511
Alfi Syahra Ginting	24,36364	40,14736	24,36364	539,5869
Amista Ginting	80,21818	27,6211	27,6211	726,9251
Anas Tasya	66,21818	15,8526	15,8526	251,3049
Angel Tridhany	16,21818	38,37894	16,21818	263,0293
Anggi Ardian	44,21818	25,49474	25,49474	649,9817
Anisa Ginting	26,78182	79,37894	26,78182	717,2688
Ardiani Sembiring	17,4909	42,37894	17,4909	305,9315
Arjuna Ginting	28,2	29,37894	28,2	795,24
Aulia Purba	16.5636	49,37894	16.5636	274,3528
Aurelia Cinta Uli	48,21818	4,73684	4,73684	22,4376

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 6 ; Nomor 1 ; Januari 2023 ; Page 235-245
E-ISSN : 2615-5133 ; **P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

Bastanta Silangit	23.29092	47.37894	23.29092	542,4669
Brando Sirait	26,8	52.37894	26,8	718,24
Brian Depari	64.2182	29.37894	29.37894	863,1221
Bripora Tarigan	35.05454	25.6	25.6	655,36
Caca Yustika	23.63636	52,37894	23.63636	558,6775
Cesylia Tamba	27.12726	38,37894	27.12726	735.8882
Chika Trivena Marbun	21.29092	39,37894	21.29092	453.3032
ClaraBenedita Sinuhaji	49,12726	27,36842	27,36842	749.0304
Claudia Silitonga	39,21818	13,37894	13,37894	178.9960
Crisda Stephani	89,21818	36,6211	36,6211	1341.1049
Dandi Lucky Cristoper	47,21818	11,3368	11,3368	128.5230
Dear Meliani Tarigan	37,21818	17,37894	17,37894	302.0275
Debora Tania	93,21818	40,6211	40,6211	1650.0737
Diego Karo	12,85454	65,37894	12,85454	165.2391
Dira Aulia	96,21818	45,7263	45,7263	2090.8945
Divia Tesalonika T	23,14546	33,37894	23,14546	535.7123
Edi Suranta Tarigan	42,21818	16,4421	16,4421	270.3426
Egi Depari	27,12726	30,37894	27,12726	735.8882
Elfransius Sembiring	80,21818	29,7263	29,7263	883.6529
Elpa Tarigan Girsang	30,87274	70,37894	30,87274	953.1260
Emelia Teodora	65,21818	24,9579	24,9579	622,8967
Emiya Pepayosa Giting	120,2182	67,6211	67,6211	4572,6131
Eriska Perangin Angin	38,21818	14,37894	14,37894	206,7539
Etisa Sinulaki	44,21818	30,27368	30,27368	916,4957
Evrani Tabina	36,05454	24,37894	24,37894	594,3327
Febry Yhana Laia	16,78182	69,37894	16,78182	281,6294
Febyola Tarigan	44,2182	28,27368	28,27368	799,4009
Finka Sihombing	81,2182	33,4	33,4	1115,56
Friska Natalia	27,63636	44,37894	27,63636	763,7683
Gebita Putri Melinia	51,05454	45,89474	45,89474	2106,3271
Geby Marselina	64,2182	13,3263	13,3263	177,5902
Gia Kinarian	31,78182	84,37894	31,78182	1010,0840
Gilang Saragih	19,8	41,37894	19,8	392,04
Gobin Ginting	49,72726	44,37894	44,37894	1969,4903
Grace Marpaung	33,4909	26,37894	26,37894	695,8484
Hafifah Cairyah	86,2182	35,7263	35,7263	1276,3685
Herdianta Tarigan	93,2182	40,6211	40,6211	1650,0737
Pebiola Br Sembiring	22,2909	32,37894	22,2909	337,7854
Petrus Siagian	34,2182	18,37894	18,37894	352,7567
Piere Traigan	18,78182	71,37894	18,78182	644,0905
Priadi Sihombing	27,2182	25,37894	25,37894	223,3667
Prilly Dwi Gita	30.36364	54,37894	30.36364	921,9506
Purnama Sari	20,78182	73,37894	20,78182	392,04
Rama Dania	19,8	47,37894	19,8	876,16
… dst				
Zidane Sion Ruben	44,2182	14,37894	14,37894	206,7539
	WCV			**115093,5334**

3. Menghitung Nilai BCV (*Between Cluster Variation*)

Karena *centroid* m1(86,03636; 85,92727; 86,34545; 86,78182; 87,58182; 82,96364), m2(73,61053; 75,14737; 77,03158; 78,88421; 81; 76,94737)

D(m1,m2) $=\sqrt{((86{,}03636-73{,}61053)^2+(85{,}92727-75{,}14737)^2+(86{,}34545-77{,}03158)^2+(86{,}78182-78{,}88421)^2+(87{,}58182-81)^2+(82{,}96364-76{,}94737)^2)}$

$= 53{,}0153$

BCV $= 53{,}0153$

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 6 ; Nomor 1 ; Januari 2023 ; Page 235-245
E-ISSN : 2615-5133 ; **P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

4. Menghitung Besar Rasio
   Rasio besaran antara BCV (*Between Cluster Variation*) dengan WCV (*Within Cluster Variation*)
   Besar Rasio = BCV/WCV
   = 53,0153/115093,5334
   = 0,00046

Karena besar rasio pada iterasi ini dengan iterasi sebelumnya masing berberda atau berubah, maka dilanjutkan ke iterasi ke 3 sesuai pada tabel 4.

Tabel 4. Hasil Perhitungan Iterasi 3

Nama	Nilai Jarak		Jarak Terdekat	Jarak Terdekat^2
	C1	C2		
Ade Depari	21,17742	49.8161	21,17742	448.4831
Aldi Ahmad Dhani	46,17742	62.74712	46,17742	2,132.3541
Aldo Girsang	76,3387	24.0345	24.0345	577.6571
Alfi Syahra Ginting	24,4355	41,14942	24,4355	597.0936
Amista Ginting	79,3387	24.1839	24.1839	584.8610
Anas Tasya	65,3387	14,9885	14,9885	224.6551
Angel Tridhany	16,5645	41,8161	16,5645	274.3826
Anggi Ardian	43,3387	27.03448	27.03448	730.8631
Anisa Ginting	27,66128	82.8161	27,66128	765.1464
Ardiani Sembiring	15,371	45.8161	15,371	236.2676
Arjuna Ginting	26,8548	32.8161	26,8548	721.1802
Aulia Purba	15,8226	52.8161	15,8226	250.3546
Aurelia Cinta Uli	47,3387	7.8161	7.8161	61.0914
Bastanta Silangit	24,6613	49.5977	24,6613	608.1797
Brando Sirait	25,371	55.8161	25,371	643.6876
Brian Depari	63,3387	28.2529	28.2529	798.2263
Bripora Tarigan	33,8548	27.88506	27.88506	777.5765
Caca Yustika	21,8871	55.8161	21,8871	479.0451
Cesylia Tamba	26,3065	41.8161	26,3065	692.0319
Chika Trivena Marbun	22,6613	42.8161	22,6613	513.5345
ClaraBenedita Sinuhaji	48,3065	28.43678	28.43678	808.6504
Claudia Silitonga	38,3387	16.8161	16.8161	282.7812
Crisda Stephani	88,3387	33.1839	33.1839	1,101.1712
Dandi Lucky Cristoper	46,3387	12.74712	12.74712	162.4890
Dear Meliani Tarigan	36,3387	19,5977	19,5977	384.0698
Debora Tania	92,3387	37.1839	37.1839	1,382.6424
Diego Karo	13,66128	68.8161	13,66128	186.6305
Dira Aulia	95,3387	43.0345	43.0345	1,851.9681
Divia Tesalonika T	23,5645	36.8161	23,5645	555.2856
Edi Suranta Tarigan	41,3387	18.45978	18.45978	340.7634
Egi Depari	26,3065	33.8161	26,3065	692.0319
Elfransius Sembiring	79.3387	27.0345	27.0345	730.8641
Elpa Tarigan Girsang	31,69353	73.8161	31,69353	1,004.4798
Emelia Teodora	64.3387	23.4828	23.4828	551.4418
Emiya Pepayosa Giting	119.339	64.1839	64.1839	4,119.5730
Eriska Perangin Angin	37.3387	17.8161	17.8161	317.4134
Etisa Sinulaki	43.3387	31.74712	31.74712	1,007.8796
Evrani Tabina	34.8548	26.88506	26.88506	722.8064
Febry Yhana Laia	17.66128	72.8161	17.66128	311.9208
Febyola Tarigan	43.3387	30.96552	30.96552	958.8634
Finka Sihombing	80.3387	31.1149	31.1149	968.1370
Friska Natalia	25.8871	47.05748	25.8871	670.1419
Gebita Putri Melinia	49.8548	47.05748	47.05748	2,214.4064
Geby Marselina	63.3387	12.3678	12.3678	152.9624

Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD
Volume 6 ; Nomor 1 ; Januari 2023 ; Page 235-245
E-ISSN : 2615-5133 ; **P-ISSN** : 2621-8976
https://ojs.trigunadharma.ac.id/index.php/jsk/index

Gia Kinarian	32.66128	87.8161	32.66128	1,066.7592
Gilang Saragih	18.371	44.8161	18.371	337.4936
Gobin Ginting	49.3065	46.05748	46.05748	2,121.2914
Grace Marpaung	32,371	29.8161	29.8161	888.9998
Hafifah Cairyah	85.3387	33.0345	33.0345	1,091.2781
Herdianta Tarigan	92.3387	37.1839	37.1839	1,382.6424
Pebiola Br Sembiring	21.7903	35.8161	21.7903	474.8171
Petrus Siagian	33.3387	21.8161	21.8161	475.9422
Piere Traigan	19.66128	74.8161	19.66128	386.5659
Priadi Sihombing	26.3387	28.8161	26.3387	693.7271
Prilly Dwi Gita	32.11292	56.5977	32.11292	1,031.2396
Purnama Sari	21.66128	76.8161	21.66128	469.2110
Rama Dania	21,62904	49.5977	21,62904	467.8153
Rani Lestari Sitepu	31.8548	31,88506	31.8548	1,014.7282
… dst				
Zidane Sion Ruben	43.3387	16.5977	16.01136	275.4836
	WCV			**117209.628**

1. Menghitung nilai BCV (*Between Cluster Variation*)
   Karena *centroid* m1 (85,3871; 85,45161; 86,24194; 86,74194; 87,74194; 82,77419), m2(73,03348; 74,58621; 76,32184; 78,27686; 80,3908; 76,57471)
   D(m1,m2) =
   $$\sqrt{(85{,}3871 - 73{,}03348)^2 + (85{,}45161 - 74{,}58621)^2 + (86{,}24194 - 76{,}32184)^2 + (86{,}74194 - 78{,}27686)^2 + (87{,}74194 - 80{,}3908)^2 + (82{,}77419 - 76{,}57471)^2}$$
   = 55,1779
   BCV = 55,1779
2. Menghitung Besar Rasio
   Rasio antara BCV (*Between Cluster Variation*) dengan WCV (*Within Cluster Variation*):
   Besar Rasio = BCV/WCV
   = 55,1179/117209,628
   = 0,00046
3. Hasil
   Karena besar rasio pada iterasi baru dengan iterasi sebelumnya sudah sama dan tidak berubah, maka iterasi dihentikan. Hasil dari peng-*claster*-an dapat dilihat pada tabel 5 sebagai berikut:

Tabel 5. Hasil Proses *K-Means*

Nama	C1	C2
Ade Depari	IPA	
Aldi Ahmad Dhani	IPA	
Aldo Girsang		IPS
Alfi Syahra Ginting	IPA	
Amista Ginting		IPS
Anas Tasya		IPS
Angel Tridhany	IPA	
Anggi Ardian		IPS
Anisa Ginting	IPA	
Ardiani Sembiring	IPA	
Arjuna Ginting	IPA	
Aulia Purba	IPA	
Aurelia Cinta Uli		IPS
Bastanta Silangit	IPA	
Brando Sirait	IPA	
Brian Depari		IPS
Bripora Tarigan		IPS
Caca Yustika	IPA	

Cesylia Tamba	IPA	
Chika Trivena Marbun	IPA	
ClaraBenedita Sinuhaji		IPS
Claudia Silitonga		IPS
Crisda Stephani		IPS
Dandi Lucky Cristoper		IPS
Dear Meliani Tarigan		IPS
Debora Tania		IPS
Diego Karo	IPA	
Dira Aulia		IPS
Divia Tesalonika T	IPA	
Edi Suranta Tarigan		IPS
Egi Depari	IPA	
Elfransius Sembiring		IPS
Elpa Tarigan Girsang	IPA	
Emelia Teodora		IPS
Emiya Pepayosa Giting		IPS
Eriska Perangin Angin		IPS
Etisa Sinulaki		IPS
Evrani Tabina		IPS
Febry Yhana Laia	IPA	
Febyola Tarigan		IPS
Finka Sihombing		IPS
Friska Natalia	IPA	
Gebita Putri Melinia		IPS
Geby Marselina		IPS
Gia Kinarian	IPA	
Gilang Saragih		IPS
Gobin Ginting		IPS
Grace Marpaung		IPS
Hafifah Cairyah		IPS
Herdianta Tarigan		IPS
Pebiola Br Sembiring	IPA	
Petrus Siagian		IPS
Piere Traigan	IPA	
Priadi Sihombing	IPA	
Prilly Dwi Gita	IPA	
Purnama Sari	IPA	
Rama Dania	IPA	
Rani Lestari Sitepu	IPA	
Regina Larasati		IPS
Renova	IPA	
Risa Natalia	IPA	
Risky Dermawan L	IPA	
Rory Surbakti	IPA	
Roy Situmorang		IPS
Ryan Giovani	IPA	
Ryan Munte	IPA	
Ryco Agustian	IPA	
Sania Engelika		IPS
Sartika Anna Tarigan		IPS
Sella Syahputi	IPA	
Selli Syahputri	IPA	
Selva Regina		IPS

Sinta Veronika P	IPA	
Surya Tarigan		IPS
Tagar Ginting		IPS
Tetti Pintaloka		IPS
Tiara Agustianda	IPA	
Tora Purba		IPS
Triganilova Ginting	IPA	
Uli Siahaan		IPS
Vioni Barus		IPS
Virgo Barus	IPA	
Virgo Ginting		IPS
Vonda Sri Karina		IPS
Wardi Perangin angin	IPA	
Wenny Lidia Kandau	IPA	
Wilmanto Siburian	IPA	
Yabest Purba		IPS
Yemima Ginting	IPA	
Yoga Syahputra		IPS
Yohanna Clarita Hts	IPA	
Yohanna Regina	IPA	
Yoki Pernando	IPA	
Yosi Monica Hutasoit		IPS
Yosua Sitanggang		IPS
Zidane Sion Ruben		IPS

4. KESIMPULAN

Berdasarkan hasil analisa dari permasalahan yang terjadi dengan kasus yang di bahas tentang rekomendasi jurusan pada SMA Negeri 2 Kabanjahe dengan menerapakan algoritma *K-means* terhadap sistem yang dirancang dan dibangun maka dapat ditarik kesimpulan sebagai berikut. Dengan menggunakan algoritma *K-means* dapat mempercepat dalam pengelompokkan data dengan menggunakan nilai *centroid* dan mencari nilai terdekat untuk mendapatkan pengelompokkan data yang menggunakan sistem yang dibuat berbasis *desktop*. Aplikasi yang dibangun dapat mempermudah siswa dalam mengambil keputusan dalam memilih jurusan berdasarkan nilai akademik dan mempermudah dalam pembuatan laporan dan menggunakan bahasa pemodela UML terdiri dari *use case diagram*, *activity diagram* dan *class diagram*. Aplikasi yang dilakukan pengujian sistem dengan menggunakan bahasa pemograman *desktop* dan menampilkan hasil dalam bentuk laporan yang menerapkan metode *K-means*.

DAFTAR PUSTAKA

[1] R. Setiawan, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik Lp3i Jakarta)," *J. Lentera Ict*, vol. 3, no. 1, pp. 76–92, 2016.

[2] S. Susanti, D. A. Irawati, and R. Rismanto, "Sistem Pendukung Keputusan Penjurusan Pada Siswa Sma," *J. Inform. Polinema*, vol. 3, no. 4, p. 48, 2017, doi: 10.33795/jip.v3i4.43.

[3] A. R. Kadafi, "Perbandingan Algoritma Klasifikasi Untuk Penjurusan Siswa SMA," *J. ELTIKOM*, vol. 2, no. 2, pp. 67–77, 2018, doi: 10.31961/eltikom.v2i2.86.

[4] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, "COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm," *J. Phys. Conf. Ser.*, vol. 1783, no. 1, pp. 1–6, 2021, doi: 10.1088/1742-6596/1783/1/012027.

[5] H. Haviluddin, S. J. Patandianan, G. M. Putra, N. Puspitasari, and H. S. Pakpahan, "Implementasi Metode K-Means Untuk Pengelompokkan Rekomendasi Tugas Akhir," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 16, no. 1, p. 13, 2021, doi: 10.30872/jim.v16i1.5182.

[6] Y. D. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.

[7] M. L. Sibuea and A. Safta, "Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustring," *Jurteksi*, vol. 4, no. 1, pp. 85–92, 2017, doi: 10.33330/jurteksi.v4i1.28.

[8] I. Vhallah, S. Sumijan, and J. Santony, "Pengelompokan Mahasiswa Potensial Drop Out Menggunakan Metode Clustering K-Means," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 2, pp. 572–577, 2018, doi: 10.29207/resti.v2i2.308.

[9] A. H. Nasyuha *et al*., "Frequent pattern growth algorithm for maximizing display items," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 19, no. 2, pp. 390–396, 2021, doi: 10.12928/TELKOMNIKA.v19i2.16192.

[10] I. Parlina, A. P. Windarto, A. Wanto, and M. R. Lubis, "Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center," *Memanfaatkan Algoritm. K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Cent. Untuk Clust. Progr. Sdp*, vol. 3, no. 1, pp. 87–93, 2018.

[11] T. Syahputra, J. Halim, and E. P. Sintho, "Penerapan Data Mining Dalam Menentukan Pilihan Jurusan Bidang Studi SMA Menggunakan Metode," *Penerapan Data Min. dalam Menentukan Pilihan Jur. di Bid. Stud. SMA menggunakan Metod. Clust. Dengan Tek. Single Link. JURTEKSI*, vol. IV, no. 2, pp. 1–4, 2018.

[12] J. Hutagalung and F. Sonata, "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah Asuransi," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 3, pp. 1187–1194, 2021, doi: 10.30865/mib.v5i3.3113.