CLUSTERING DATA UNTUK REKOMENDASI PENENTUAN JURUSAN PERGURUAN TINGGI MENGGUNAKAN METODE K-MEANS

Pareza Alam Jusia[1], Fadhel Muhammad Irfan[2], Kurniabudi[3]
[1,2,3]STIKOM Dinamika Bangsa Jambi
Jl. Jend. Sudirman Thehok Jambi
E-mail : parezaalam@gmail.com [1], fadhelmirfan23@gmail.com [2], kbudiz@yahoo.com [3]

ABSTRAK

Siswa-siswi SMA Negeri 2 Kota Jambi cenderung memilih jurusan berdasarkan karena minat, dan keinginan orang tua. Beberapa di antaranya sudah memperhitungkan potensi yang ada pada diri mereka, maka komitmen untuk belajar dibidang itu tidak akan berjalan lancar, padahal jurusan yang dia pilih itu tidak sesuai kemampuannya. Oleh karena itu, penulis melakukan analisis data *mining* menggunakan data nilai siswa kelas XII dari semester satu sampai empat dan kuisoner yang penulis bagikan. Dalam melakukan analisis penulis menggunakan alat bantu *tools WEKA* dan *RapidMiner*. Metode yang digunakan adalah metode *k-means clustering* dengan 24 atribut dan 5 *cluster*. Jumlah *cluster* pada perhitungan manual adalah, C1 terdapat 62 data, C2 terdapat 28 data, C3 terdapat 30 data, C4 terdapat 30 data, C5 terdapat 60 data. Jumlah *cluster* pada perhitungan *RapidMiner* adalah, C1 terdapat 35 data, C2 terdapat 55 data, C3 terdapat 58 data, C4 terdapat 35 data, C5 terdapat 27 data. Jumlah *cluster* pada perhitungan *WEKA* adalah, C1 terdapat 30 data, C2 terdapat 49 data, C3 terdapat 41 data, C4 terdapat 32 data, C5 terdapat 58 data.

.

Kata kunci : ***Data Mining, K-Means, Clustering, WEKA, RapidMiner, SMA.***

ABSTRACT

The students of SMA Negeri 2 city of Jambi tend to choose majors based on interest, and desire because of parents. Some of them already take into account the existing potential in them, then commitment to learning in the field of it won't go smoothly, even though the Department he chooses it doesn't match his ability. Therefore, the author does analysis of data mining using value data class XII students from one to four semesters and kuisoner the authors share. In doing the analysis the author using tools tools WEKA and RapidMiner. The method used is the method of k-means clustering with 24 attributes and 5 clusters. The number of clusters on a manual calculation is, there are 62 C1, C2 data there are 28 data, data, there are 30 C3 C4 C5 there are 30 data, there are 60 data. The number of clusters in the calculation of RapidMiner is there are 35, C1, C2 data there are 55 data, there are 58 C3 data, there are 35 C4 C5, there are data 27 data. The number of clusters on a calculation of the WEKA is a, C1, C2 data there are 30 there are 49 data, there are 41 data, the C3 C4 C5 32 there are data, there are 58 data.

Keyword : ***Data Mining, K-Means, Clustering, WEKA, RapidMiner, SMA.***

1. PENDAHULUAN

Penentuan jurusan akan berdampak terhadap kegiatan akademik selanjutnya dan mempengaruhi pemilihan bidang ilmu atau studi bagi siswa-siswi yang ingin melanjutkan ke perguruan tinggi nantinya. Penentuan jurusan yang dilakukan selama ini mempunyai banyak kelemahan, antara lain berdasarkan keinginan siswa tanpa melihat latar belakang nilai akademisnya. Sehingga jurusan yang dipilih terkadang menjadi masalah bagi siswa di kemudian hari, sebagai contoh nilai akademik yang tidak maksimal, pemilihan program studi saat melanjutkan ke jenjang perguruan tinggi yang terkendala akibat jurusan SMA yang tidak sesuai, dan lain-lain.

Berdasarkan hasil wawancara dengan wakil kesiswaan Dwi Wahyuningsih, M.Pd.,kons mengatakan, siswa-siswi SMA Negeri 2 Kota Jambi cenderung memilih jurusan berdasarkan karena minat, dan keinginan orang tua. Beberapa di antaranya sudah memperhitungkan potensi yang ada pada diri mereka, maka komitmen untuk belajar dibidang itu tidak akan berjalan lancar, padahal jurusan yang dia pilih itu tidak sesuai kemampuannya. Harapannya pihak sekolah bisa melihat persentase keakuratannya untuk penentuan jurusan Perguruan Tinggi Negeri pada siswa-siswi SMA Negeri 2 Kota Jambi, jika metode ini berhasil dan persentasenya tinggi, pihak sekolah akan menggunakan kembali metode ini untuk merekomendasikan jurusan Perguruan Tinggi Negeri untuk siswa-siswi selanjutnya.

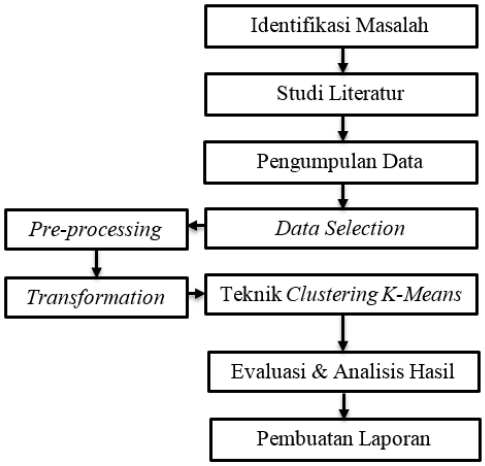
Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. tujuan utama *data mining* adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki (Jiawei Han, Micheline Kamber, 2011).

Teknologi *clustering* data merupakan suatu teknik yang menunjukan persamaan karakteristik dalam suatu kelompok sehingga akan menghasilkan informasi yang bermaanfaat. Algoritma *clustering* data sudah banyak dipergunakan diberbagai bidang misalnya untuk proses pengolahan citra, data mining proses pengambilan keputusan, pengenalan pola, maupun dalam bidang bioinformatika (Jusia, 2016). Ada beberapa algoritma yang untuk dapat melakukan proses *clustering* pada suatu dataset dalam jumlah yang banyak. Pada penelitian ini, peneliti akan menggunakan metode algoritma *K-Means* dalam menentukan jumlah cluster terbaik. *K-Means* merupakan algoritma yang sangat banyak dipergunakan karena effektif dan effisien. Ini dikarenakan *K-means* sangat mudah dipelajari dan dari segi waktu proses komputasinya relatif singkat (Jusia, 2018). Untuk itu penulis menggunakan metode *k-means clustering* bagaimana merekomendasi penentuan memilih jurusan di perguruan tinggi setelah lulus SMA dengan atribut yang digunakan diantaranya yaitu hobi, minat, bakat, sifat, dan nilai rata-rata dari mata pelajaran siswa-siswi, yaitu : Pendidikan Agama dan Budi Pekerti, Pendidikan Pancasila dan Kewarganegaraan, Bahasa Indonesia, Matematika, Sejarah Indonesia, Bahasa Ingris, Seni Budaya, Pendidikan Jasmani Olahraga dan Kesehatan, Prakarya dan Kewirausahaan, Fisika IPA, Matematika Peminatan IPA, Kimia IPA, Biologi IPA, Ekonomi IPA, Bahasa Inggris IPA, Ekonomi IPS, Sosiologi IPS, Sejarah IPS Geografi IPS, Bahasa Inggris IPS

2. METODOLOGI

Kerangka kerja penelitian merupakan tahapan-tahapan yang dilakukan selama mengerjakan penelitian. Kerangka kerja penelitian dibuat agar

mempermudah pencapaian hasil penelitian, dapat menyelesaikan penelitian tepat waktu dan penelitian dapat berjalan sesuai dengan yang diharapkan. Adapun kerangka kerja penelitian yang digunakan dapat dilihat pada gambar.



Gambar 1. Kerangka Kerja Penelitian

3. LANDASAN TEORI

3.1 Clustering

Clustering adalah proses pengelompokan kumpulan data menjadi beberapa kelompok sehingga objek di dalam satu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaan dengan objek dikelompok lain. Perbedaan dan persamaannya biasanya berdaasarkan nilai atribut dari objek tersebut dan dapat juga berupa perhitungan jarak. *Clustering* sendiri juga disebut *unsupervised classification*, karena *clustering* lebih bersifat untuk dipelajarai dengan diperhatikan. *Cluster analysis* merupakan proses partisi satu set objek data ke dalam himpunan bagian. Setiap himpunan bagian adalah *cluster*, sehingga objek yang ada di dalam cluster mirip satu sama dengan lainnya, dan mempunyai perbedaan dengan objek dari *cluster* yang lain. Partisi tidak dilakukan dengan manual algoritma *clustering*. Oleh karena itu, *clustering* sangat berguna dan bisa menemukan grup yang tidak dikenal dalam data. *Cluster analysis* banyak digunakan dalam berbagai aplikasi seperti *Business Intelligence, Image Pattern Recognition, Web Search, Biology,* dan *Security* (Jiawei Han, Micheline Kamber, 2011). Di dalam *business intelligence, clustering* bisa mengatur banyak *customer* ke dalam banyak grup. Contohnya pengelompokan *customer* ke dalam beberapa *cluster* dengan persamaan karakteristik yang kuat. *Clustering* juga dikenal sebagai data *segmentation*, karena *clustering* mempartisi banyak *data set* ke dalam banyak grup berdasarkan persamaannya. *Clustering* juga bisa sebagai *outlier detection*, di mana *outlier* bisa menjadi menarik daripada kasus yang biasa. Aplikasinya adalah *Outlier Detection*, untuk mendeteksi *card fraud* dan memonitori aktivitas kriminal dalam *e-commerce*. Contohnya adalah pengecualian dalam transaksi kartu kredit (Florin Gorunescu, 2011).

Teknik *Clustering K-Means* merupakan algoritma *clustering* sederhana yang bersifat tanpa arahan (*unsupervised*). Misalkan *D* adalah sebuah *dataset* dari *n* objek, dan k adalah jumlah *cluster* yang akan dibentuk, algoritma partisi mengatur objek-objek tersebut ke dalam partisi *k* ($k \leq n$), di mana setiap partisi menggambarkan sebuah *cluster*. Setiap *cluster* dibentuk untuk mengoptimalkan kriteria partisi, seperti fungsi perbedaan berdasarkan jarak, sehingga objek-objek di dalam sebuah *cluster* adalah mirip, sedangkan objek-objek pada *cluster* yang berbeda adalah tidak mirip dalam hal atribut *dataset*. Persamaan untuk menghitung jarak antar data pada *K-Means* menggunakan rumus *Euclidiance Distance* (D) yang ditunjukkan pada persamaan (Larose & Larose, 2014).

$$D(x_2, x_1) = \sqrt{\sum_{j=1}^{p}(x_{2j} - x_1)^2} \quad \text{.....(1)}$$

keterangan :
p = Dimensi data
x_1 = Posisi titik 1
x_2 = Posisi titik 2

Algoritma standar dari *K-Means* adalah sebagai berikut (Larose & Larose, 2014) :

1. Tentukan jumlah *clustering* yang diinginkan (misalkan : k3).
2. Pilih centroid awal secara acak. Pada langkah ini secara acak akan dipilih 3 buah data sebagai *centroid*.
3. Hitung jarak dengan *centroid*. Pada langkah ini setiap data akan ditentukan *centroid* terdekatnya, dan data tersebut akan ditetapkan sebagai anggota kelompok yang terdekat dengan *centroid*. Untuk menghitung jarak ke *centroid* masing-masing *cluster*. Misalkan data (x,y), *centroid* M1 : (a1,b1), *centroid* M2 : (a2,b2), *centroid* M3 : (a3,b3).

$$DM1 = \sqrt{(x-a1)^2 + (y-b1)^2} = ? \quad \ldots (2)$$
$$DM2 = \sqrt{(x-a2)^2 + (y-b2)^2} = ? \quad \ldots (3)$$
$$DM3 = \sqrt{(x-a3)^2 + (y-b3)^2} = ? \quad \ldots (4)$$

Buat tabel hasil perhitungan jarak selengkapnya antara masing-masing data dengan *centroid*, maka di dapatkan keanggotaan dari masing-masing *cluster*

Pada langkah ini dihitung pula rasio antara BCV (*Between Cluster Variation*) dengan WCV (*Within Cluster Variation*) :

Karena *centroid* M1 = (a1,b1), M2 = (a2,b2), M3 = (a3,b3)

$$d(m1,m2) = \sqrt{(a1-a2)^2 + (b1-b2)^2} = ? \quad \ldots . (5)$$

$$d(m1,m3) = \sqrt{(a1-a3)^2 + (b1-b3)^2} = ? \quad \ldots . (6)$$

d(m2,m3)= BCV = d(m1,m2) + d(m1,m3) + d(m2,m3) = ?

Dalam hal ini d(mi,mj) menyatakan jarak *euclidean* dari m ke mj. Menghitung WCV yaitu dengan memilih jarak terkecil yang terdapat pada tabel keanggotaan.

$WCV = c1^2+c2^2+c3^2+ N = ?$

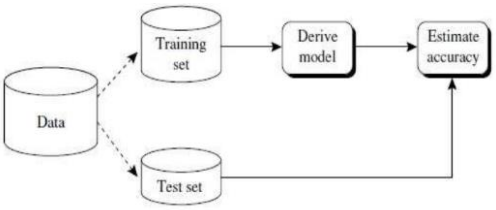
Sehingga besar rasio = BCV/WCV = ?
Karena langkah ini merupakan iterasi 1 maka lanjukan ke langkah berikutnya.

4. Pembaruan *centroid* dengan menghitung rata-rata nilai pada masing-masing *cluster*. Setelah menghitung rata-rata nilai pada masing-masing *cluster* didapatkan *centroid* baru yaitu :
   M1 = (a1,b1), M2 = (a2,b2), M3 = (a3,b3).
5. Iterasi ke 2 kembali kelangkah ke 3, jika masih ada data yang berpindah *cluster* atau jika nilai *centroid* diatas ambang, atau jika pada fungsi obyektif yang digunakan masih diatas ambang. Jika tidak maka iterasi dihentikan.

3.2 Data training & data testing

Dalam klasifikasi data pada umumnya dibagi menjadi dua, yaitu *data training* dan *data testing*. Untuk membentuk sebuah model klasifikasi, maka dilakukan *data training* yang mana *data training* tersebut biasanya digunakan oleh algoritma klasifikasi (misalnya *decision tree, bayesian, neural network, SVM*) (Jusia, 2017). Representasi pengetahuan dari model yang telah dihasilkan tersebut dapat digunakan untuk mengukur sejauh mana tingkat keberhasilan dari klasifikasi tersebut melakukan klasifikasi dengan benar. Oleh karena itu, pada saat melakukan *testing*, data yang diuji seharusnya tidak terdapat pada data *training*, sehingga dapat diketahui apakah model klasifikasi dapat melakukan klasifikasinya dengan baik. Proporsi untuk melakukan data *training* biasanya lebih besar dibanding data *testing* dan biasanya 2/3 dari total data dijadikan sebagai data *training*, sedangkan sisanya akan digunakan sebagai data *testing* inilah yang disebut dengan *holdout method*. Menurut (Jiawei Han, Micheline Kamber, 2011), *holdout method* adalah data yang diberikan secara acak dibagi menjadi dua set independen, yaitu *training set* dan *test set*. Biasanya, dua pertiga dari data yang dialokasikan untuk *training set*, dan sisanya, sepertiga

dialokasikan untuk test set. *Training set* digunakan untuk menurunkan model. Akurasi model tersebut kemudian diperkirakan dengan *test set*.



Gambar 2. Holdout Method (Jiawei Han, Micheline Kamber, 2011)

3.3 Confusion matrix

Confusion matrix digunakan ntuk mengevaluasi kinerja dari suatu metode atau model, maka diperlukannya sebuah cara yang sistematis. Pada evaluasi klasifikasi didasarkan pengujian pada objek yang benar dan salah. Untuk menentukan jenis terbaik dari skema pembelajaran yang digunakan, maka menggunakan validasi data yang berdasarkan data pelatihan untuk melatih skema pembelajaran *Confusion matrix* berisi informasi mengenai hasil klasifikasi aktual dan yang telah diprediksi oleh sistem klasifikasi. Performa dari sistem tersebut biasanya dievaluasi menggunakan data dalam sebuah matriks. *Confusion matix* juga merupakan tabel yang digunakan sebagai alat ukur yang berguna untuk melakukan analisis seberapa baik hasil pengklasifikasian yang benar dan salah dari hasil prediksi yang telah dilakukan dalam kelas yang berbeda-beda (Florin Gorunescu, 2011).

3.4 K-fold cross validation

K-fold cross validation merupakan salah satu teknik untuk melakukan estimasi tingkat kesalahan pengujian pemrosesan citra digital. Cara kerja *K-fold cross validation* yaitu dengan mengelompokkan data latih dan data uji yang saling terpisah, kemudian melakukan proses pengujian yang diulang sebanyak *K* kali (Florin Gorunescu, 2011). Langkah dari *K-fold cross validation* antara lain : (1) Membagi data asli yang tersedia menjadi *K* kelompok; (2) Setiap *K* dibuat sejumlah T himpunan data yang memuat semua data latih kecuali yang berada di kelompok ke-k; (3) Mengerjakan algoritma yang dimiliki dengan sejumlah T data latih; (4) Pengujian algoritma menggunakan data pada kelompok *K* sebagai data uji; (5) Melakukan pencatatan hasil algoritma (Quinlan, 1999). Keuntungan dari teknik *K-fold cross validation* ini yaitu menunjukkan bahwa semua elemen pada baris data digunakan untuk pelatihan sekaligus pengujian.

4. HASIL DAN PEMBAHASAN

4.1 Representasi data

Berdasarkan hasil dari wawancara dan pembagian kuesioner yang sudah dilakukan serta data nilai semester 1 sampai 4 yang dirata-ratakan, penulis memperoleh data-data siswa SMA Negeri 2 Kota Jambi. Jumlah seluruh siswa di SMA Negeri 2 Kota Jambi ada 395 siswa yang terdiri 11 kelas, 6 kelas IPA dan 5 kelas IPS. Dikarenakan setiap kelas XII ada yang berjumlah 19 sampai 30 siswa, maka penulis mengambil sampel 19 siswa untuk 10 kelas, dan 20 siswa untuk 1 kelas, sehingga jumlah keseluruhan siswa dari 11 kelas yaitu 210 siswa. Atribut yang digunakan pada seluruh data siswa SMA Negeri 2 Kota Jambi berjumlah 24, yaitu Hobi, Minat, Bakat, Sifat, nilai rata-rata mata pelajaran untuk jurusan IPA, dan nilai rata-rata mata pelajaran untuk jurusan IPS. Atribut tersebut dipilih penulis karena penentuan jurusan dicari berdasarkan nilai semester 1 sampai 4, dan hasil kuesioner siswa, setellah itu data terserbut diolah manual menggunakan *clustering k-means*

4.2 Transformasi data

Agar data di atas dapat diolah dengan menggunakan metode *k-means clustering*, maka data yang berjenis data nominal seperti hobi, minat, bakat, sifat harus diinisialisasikan terlebih dahulu dalam

bentuk angka. Berikut inisialisasi tiap atribut:

Tabel 1. Inisialisasi Hobi

No.	Keteragngan Hobi	Inisial
1.	Membaca	1
	Menulis	2
	Menghitung	3
	Traveling	4
	Memasak	5
	Olahraga	6
	Menyanyi	7
	Melukis/menggambar	8
	Photography	9
	Fashion	10

4.3 Perhitungan *K-Means Clustering*

Perhitungan dilakukan dengan menggunakan Persamaan untuk menghitung jarak antar data pada *K-Means* menggunakan rumus *Euclidiance Distance (D)* yang ditunjukkan pada persamaan rumus (1).

1. Tentukan jumlah cluster yang diinginkan (cluster = 5).
2. Pilih centoid awal secara acak. Pada langkah ini secara acak akan dipilih 5 buah data sebagai centroid, data {40,80, 120, 160, 200}.
3. Hitung jarak dengan centroid (iterasi 1)

   Pada langkah ini setiap data akan ditentukan centroid terdekatnya, dan data tersebut akan diterapkan sebagai anggota kelompok yang terdekat dengan centroid. Untuk menghitung jarak ke centroid masing-masing cluster pada siswa/siswi no. 1 sebagai berikut :

Data yang digunakan : (3, 8, 3, 10, 84, 85 , 86, 81, 90, 78, 86, 81, 79, 84, 77, 87, 83, 81, 0, 0, 0, 0, 0, 0, 0),

Centroid M1 : (2, 3, 10, 10, 78, 80, 79, 82, 84, 79, 88, 88, 80, 76, 76, 83, 84, 81, 0,0,0,0, 0, 0).

Centroid M2 : (10, 9, 4, 1, 80, 77, 81, 77, 87, 78, 81, 81, 85, 79, 76, 75, 79, 0, 78, 0, 0, 0,0,0).

Centroid M3 : (10, 4, 1, 7, 77, 79, 78, 80, 81, 76, 79, 77, 73, 79, 78, 79, 80, 0, 75, 0,0,0,0,0).

Centroid M4 : (1, 9, 4, 1, 76, 83, 78, 75, 81, 72, 83, 80, 80,0,0,0,0,0,0,74,83,81,73,80).

Centroid M5 : (4, 2, 5, 6, 73, 78, 76, 73, 75, 68, 79, 77, 77, 0,0,0,0,0,0,72, 78, 75, 69, 73).

$$DM1=\sqrt{\begin{matrix}(3-2)^2+(8-3)^2+(3-10)^2+(10-10)^2+\\(84-78)^2+(85-80)^2+(86-79)^2+(81-82)^2+\\(90-84)^2+(78-79)^2+(86-88)^2+(81-88)^2+\\(79-80)^2+(84-76)^2+(77-76)^2+(87-83)^2+\\(83-84)^2+(81-81)^2+(0-0)^2+(0-0)^2+\\(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2=19{,}01973\end{matrix}}$$

Tabel 2. perhitungan jarak antara masing-masing data dengan *centroid* (iterasi 1)

No.	Nama	PENCARIAN CLUSTER PERTAMA					JARAK TERDEKAT				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
1	Agung Satria	19,01973	114,8442	112,8162	255,0376	248,7085	✓				
2	Arlingga Alfaesi	34,02389	114,3091	110,3885	241,1964	233,1477	✓				
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	...
68	Dwi Indriyani Sari	114,2522	15,71226	15,75397	244,4331	237,3622		✓			
69	Aulira Putri Maharani	111,1201	17,5784	15,46569	244,4774	237,3819			✓		
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
80	Annisa Alsadila	114,5723	0	18,65978	246,5431	239,7417		✓			
81	Arjun Sandi Wiranata	111,7464	15,36839	16,06238	247,0466	240,1233		✓			
82	Yos Vinora	115,138	14,66927	21,03865	247,7892	241,1095		✓			
83	Nabila Maghfirah	112,2631	14,69481	17,67767	245,985	239,1748		✓			
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
95	Putri Syskina	112,3975	19,6023	18,46788	247,4511	240,5814			✓		
96	Tariski Amanah W	111,9872	25,15576	18,89444	246,3331	238,997			✓		
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
105	Arum Sah Putri	111,6915	20,29932	18,02776	249,9546	243,1615			✓		
106	Marshella Monica	119,3273	28,1436	31,95309	257,3193	251,3122		✓			
107	Tri Widodo	112,9361	23,7671	23,49867	250,0582	243,53			✓		
108	Muhammad Syaifudin	112,781	16,91523	19,64847	248,671	241,767		✓			
109	Bella Yurice	108,959	20,95233	21,26176	245,5771	238,5682		✓			
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
122	Pandhu Pramudya	246,9755	242,542	243,6242	13,61114	17,96513				✓	
123	Rina Delvianata	243,5548	239,0269	239,8236	18,86032	13,16516					✓
124	Aditya Raihan Asmara	244,1112	239,7842	240,528	17,64673	9,355455					✓
125	Selvi Oktaviani	246,5617	242,3816	243,0894	16,83243	15,90833					✓
126	Dira Ramadhanti	243,1488	238,7186	239,0822	36,17482	24,86288					✓
127	Afiq Julianayah	242,0238	237,2212	237,7348	24,61897	12,8637					✓
128	Elfin Nadila	245,3533	240,7355	241,7917	15,66262	15,38019					✓
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
138	Wulan Aprilia	245,3041	241,2572	241,953	14,75561	10,37822					✓
139	Nadilla Putri Arimbi	240,9634	236,3592	236,6528	26,35624	14,17368					✓
140	Andre Ardiansyah	242,6274	238,3523	238,6609	23,54295	10,89837					✓
141	Nurul Ismany Putri	248,8482	244,6187	245,096	16,67601	15,36328					✓
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
147	Desviano Ramadhani	245,8986	241,6874	242,3328	17,21441	15,12756					✓
148	Mami Yanti	250,669	246,7538	247,7694	10,60536	20,04543				✓	
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
156	Christia Adi	247,1013	243,1913	243,9122	16,08803	14,8182					✓
157	Yogi Kurnia	254,0839	249,997	251,3014	16,16727	26,16401				✓	
158	Dandi Cahyo	242,2359	237,6448	238,4581	19,86256	12,51058					✓
159	Seli Karyana Aryanti	246,8597	242,4718	243,3282	10,72931	11,21937				✓	
160	Caroline Zaitri	250,8904	246,5431	247,6766	0	18,44974				✓	
161	Dena Tasya Vira	247,02	242,484	243,5212	13,6454	14,17842				✓	
162	Resi Mardila	244,0975	239,3998	240,0526	21,43576	13,68035					✓
163	Hanif El-hadi	247,3934	243,2165	244,3749	13,97091	18,87671				✓	
164	Maykel Runda	240,862	236,0737	236,9844	23,16736	15,57164					✓
165	Risci Ayu Oktaviani	247,5692	243,4027	244,2136	16,78318	14,42317					✓
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
173	Benny	244,3996	240,1308	240,8626	20,79618	16,58301					✓
174	Ari Hidayat	252,4173	248,2032	249,4762	15,31649	26,61857				✓	
.....	.....	.....	.....	.....	...	.....	.....	.....	.....	.....	...
181	Miranda	247,3944	242,8606	243,7471	13,10107	13,18956				✓	
182	Jodi	239,2404	234,6012	234,7796	31,36586	18,10969					✓
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
203	Citia Kartika	246,6449	242,6912	243,2087	14,53907	12,73359					✓
204	Cindy Artika Hutabarat	242,63	238,5272	238,7526	21,83879	10,74959					✓
205	Cindy Amanda Putri	243,6066	238,9706	239,6349	20,44722	11,24879					✓
206	Dina Utami	247,949	243,2933	244,1298	16,34535	16,2868					✓
.....	.....	...	.....	...	.....	.....	.....	.....	.....	.....	.....
210	Ayu Dewi Retno	249,7239	245,6857	246,5471	13,81549	17,82799				✓	
JUMLAH							148	180	182	185	145
							62	30	28	25	65

Pada langkah ini dihitung pula rasio antara besaran *BCV (Between Cluster Variation*) dengan *WCV (Within Cluster Variation*) :

Centroid M1 : (2,3,10,10,78,80,79,82,84,79,88,88,80,76,76, 83,84,81,0,0,0,0,0,0),

Centroid M2 : (10,9,4,1,80,77,81,77,87,78,81,81,85,79,76,75,79,0,78,0,0,0,0,0),

Centroid M3 : (10,4,1,7,77,79,78,80,81,76,79,77,73,79,78,79,80,0,75,0,0,0,0,0),

Centroid M4 : (1,9,4,1,76,83,78,75,81,72,83,80,80,0,0,0,0,0, 0,74,83,81,73,80),

Centroid M5 : (4,2,5,6,73,78,76,73,75,68,79,77,77,0,0,0,0,0, 0,72,78,75,69,73)

BCV (Between Cluster Variation) =

$$d(m1,m2)=\sqrt{\begin{array}{c}(2-10)^2+(3-9)^2+(10-4)^2+(10-1)^2+\\(78-80)^2+(80-77)^2+(79-81)^2+(82-77)^2+\\(84-87)^2+(79-78)^2+(88-81)^2+(88-81)^2+\\(80-85)^2+(76-79)^2+(76-76)^2+(83-75)^2+\\(84-79)^2+(81-0)^2+(0-78)^2+(0-0)^2+\\(0-0)^2+(0-0)^2+(0-0)^2+(0-0)^2\end{array}}=114{,}5723025$$

Dalam hal ini d(mi,mj) menyatakan jarak *Euclidean* dari m ke mj. Menghitung WCV dengan memilih jarak terkecil antara data dengan *centroid* pada masing-masing *cluster*

Tabel 3. Jarak terdekat (iterasi 1)

No.	Nama	JARAK TERDEKAT	PROSES WCV
1	Agung Satria	19,01973	361,75
2	Arlingga Alfarezi	34,02389	1157,625
3	Deni Tawaf Suharta	20,87313	435,6875
4	Dinda Chairunnisa	23,5717	555,625
5	Shalza Dwi Adistia	19,84943	394
----	----	----	----
200	Dila Amalia	0	0
201	Ribka Mutiara Sinurat	9,870453	97,42585
202	Fita Oktaviani	25,74505	662,8078
203	Citia Kartika	12,73359	162,1443
204	Cindy Artika Hutabarat	10,74959	115,5537
205	Cindy Amanda Putri	11,24879	126,5352
206	Dina Utami	16,2868	265,2598
207	Ilyas Nurahman	15,79754	249,5624
208	Hary Satrya	15,3847	236,6889
209	Awie Danu Bintoro	11,80006	139,2413
210	Ayu Dewi Retno	13,81549	190,8678
NILAI WCV			72992,43

WCV=$19{,}0197266^2 + 34{,}02388867^2 + 20{,}87312866^2 + 23{,}57169913^2 + ... +0^2 + 17{,}62455673^2 + 13{,}67936402^2 + ... + 0^2 + 15{,}36839289^2 + 14{,}6692706^2 + ... + 0^2 + 12{,}30822821^2 + 13{,}61114108^2 + ... + 0^2 + 13{,}64540044^2 + 13{,}68034949^2 + ... + 0^2 + 9{,}870453187^2 + 11{,}80005529^2 + 13{,}81549041^2 = 72992{,}43$

Sehingga besar rasio = BCV/WCV = 1320,37976/72992,43 = 0,0180892689.

Karena langkah ini merupakan iterasi 1 maka lanjutkan ke langkah berikutnya.

4. Pembaharuan *centroid* dengan menghitung rata-rata nilai pada masing-masing *cluster*.

Tabel 4. Nilai *centroid* pada iterasi 2

cluster 1	4	5	6	5	83	85	84	85	87	79	86	83	81	83	79	84	84	82	0	0	0	0	0	0
cluster 2	5	6	6	4	84	86	83	84	85	79	84	81	84	81	78	80	81	0	79	0	0	0	0	0
cluster 3	5	6	4	6	77	78	77	82	79	75	81	81	77	78	75	77	77	0	74	0	0	0	0	0
cluster 4	4	6	5	4	78	82	80	78	81	75	82	81	90	0	0	0	0	0	0	75	81	81	74	77
cluster 5	5	6	6	5	72	77	76	73	76	69	78	79	74	0	0	0	0	0	0	71	75	76	71	70

Dengan langkah pengolahan data yang sama menggunakan nilai centroid baru pada Iterasi ke-2 maka di dapat hasil jarak centroid yang tertera pada tabel berikut :

Tabel 5. Perhitungan jarak antara masing-masing data dengan *centroid* (iterasi 2)

No.	Nama	PENCARIAN CLUSTER KEDUA					JARAK TERDEKAT				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
1	Agung Satria	9,924717	114,1687	112,3343	254,2223	247,553	✓				
2	Arlingga Alfarezi	38,36828	116,7267	109,5542	241,1755	231,8591	✓				
...	....	....	.....	....	....	....	....	....	....	....	...
68	Dwi Indriyani Sari	115,1029	20,02967	15,49798	244,0065	235,9782			✓		
69	Aulira Putri Maharani	111,9579	16,88379	11,98176	243,9161	236,178			✓		
....	.....	....	....	....	....	....	....	....	....	....	....
80	Annisa Alsadila	114,9535	16,24231	16,47157	245,5938	238,4739		✓			
81	Arjun Sandi Wiranata	111,9246	10,54158	14,56451	246,1496	239,0411		✓			
82	Yos Vinoza	115,0788	10,6595	19,31644	246,8676	240,0273		✓			
83	Nabila Maghfirah	112,6166	16,17096	14,71394	245,3589	237,8235			✓		
....	....	....	....	....	....	....	....	....	....	....	....
95	Putri Syekina	112,0349	15,77379	16,63768	246,6056	239,3926		✓			
96	Tarizki Amanah W	112,9773	24,97749	15,12862	245,6204	237,6644			✓		
....	....	....	....	...	....	...	....	....	...	...	....
105	Arum Sah Putri	111,4966	11,83216	16,12452	248,9719	242,0434		✓			
106	Marshella Monica	118,3675	14,67566	29,83915	256,155	250,1867		✓			
107	Tri Widodo	112,8381	15,13068	20,30856	249,2497	242,3013		✓			
108	[illegible]	[illegible]	[illegible]	15,1306	247,510	240,733		✓			
109	Bella Yurice	109,3901	16,94661	15,78567	244,5704	237,5441			✓		
...	....	....	....	....	....	....	....	....	....	....	....
122	Pandhu Pramudya	250,8583	246,5547	240,3421	16,88487	15,94081					✓
123	Riza Delviansta	247,6401	243,2664	236,6808	22,25058	9,562225					✓
124	Aditya Raihan Asmara	248,3811	244,0399	237,2942	22,01413	7,881627					✓
125	Selvi Oktaviani	250,5359	246,2352	240,0223	16,68953	16,08699					✓
126	Dira Ramadhanti	246,6727	242,3469	235,0623	29,84374	9,811165					✓
127	Afiq Juliansyah	246,2156	241,7587	234,6607	27,30698	10,51215					✓
128	Elfin Nadila	249,3458	244,9453	238,6082	16,14737	13,20065					✓
....	....	....	....	....	....	....	....	....	....	....	....
138	Wulan Aprilia	249,4719	245,1714	238,7113	17,76699	10,16528					✓
139	Nadilla Putri Arimbi	245,3043	240,8427	233,5119	29,57512	11,72418	I				✓
140	Andre Ardiansyah	246,8985	242,4936	235,5587	27,21124	10,00008					✓
141	Nurul Ismany Putri	252,788	248,5958	242,1293	20,41698	15,26311					✓
....	....	....	....	....	....	....	....	....	....	....	....
147	Desviano Ramadhani	249,9	245,5623	239,1893	18,94549	13,41671					✓
148	Marni Yanti	254,5324	250,3261	244,4835	12,68119	21,51113				✓	
....	....	....	.....	....	....	....	....	....	....	....	...
156	Christia Adi	251,1677	246,8785	240,7146	17,3552	16,33342					✓

Dari tabel 9 didapatkan keanggotaan siswa/siswi SMA Negeri 2 Kota Jambi (iterasi 2), terjadi perubahan pada No. 68, 83,95,105,107,109,119, yang pada awalnya berada di *cluster* 2 berpindah pada *cluster* 3. Terjadi perubahan pada No. 122, 159,161, 181, yang pada awalnya berada di *cluster* 4 berpindah ke *cluster* 5. Selanjutnya, dihitung pula rasio antara besaran BCV (Between Cluster Variation) dengan WCV (Within Cluster Variation) pada iterasi ke-2 dengan cara pengolahan yang sama pada proses awal dengan demikian tercatat nilai BCV, WCV dan RATIO pada proses awal dan iterasi 1 seperti pada tabel dibawah ini :

Tabel 6. Perbandingan *BCV,WCV* dan *ratio* iterasi ke 1 dan 2

Nilai	Iterasi 1	Iterasi 2
BCV	1320,38	1338,572
WCV	72992,43	43181,4
Rasio	0,018089	0,030999

Melihat tabel perbandingan diatas didapat informasi karena ada data yang berpidah cluster, serta nilai ratio pada iterasi ke 2 lebih besar dari ratio pada iterasi pertama, maka iterasi dilanjutkan ke iterasi selanjutnya :

Tabel 7. Perbandingan *BCV, WCV* dan *ratio* iterasi ke 1,2,3,4,5, dan 6

Nilai	Iterasi 1	Iterasi 2	Iterasi 3	Iterasi 4	Iterasi 5	Iterasi 6
BCV	1320,38	1338,572	1335,756	1335,717	1335,443	1335,43
WCV	72992,43	43181,4	40430,36	39999,09	39916,32	39887,36
Rasio	0,018089	0,030999	0,033038	0,033394	0,033456	0,03348

Dengan langkah yang sama seperti pada iterasi sebelumnya, maka hasil pengolahan data pada iterasi ke 7 adalah sebagai berikut :

Tabel 8. Nilai *centroid* pada iterasi ke 7

cluster 1	4	5	6	5	83	85	84	85	87	79	86	83	81	83	79	84	84	82	0	0	0	0	0	0
cluster 2	5	5	6	5	85	87	84	86	85	79	84	81	84	81	78	81	82	0	79	0	0	0	0	0
cluster 3	5	6	5	5	77	78	77	80	79	75	80	81	78	78	74	77	77	0	74	0	0	0	0	0
cluster 4	5	6	5	5	78	81	79	78	80	74	82	81	79	0	0	0	0	0	0	75	80	80	74	76
cluster 5	4	6	6	4	71	77	75	73	76	68	77	78	74	0	0	0	0	0	0	71	75	75	70	69

Dengan langkah pengolahan data yang sama menggunakan nilai centroid baru pada Iterasi ke-8 maka di dapat hasil jarak centroid yang tertera pada tabel berikut :

Tabel 9. perhitungan jarak antara masing masing data dengan *centroid* (iterasi 7)

No.	Nama	PENCARIAN CLUSTER KE DELAPAN					JARAK TERDEKAT				
		C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
1	Agung Satria	9,924717	114,1468	112,4389	253,1344	246,8724	✓				
2	Arlingga Alfarezi	38,36828	117,3888	109,4469	239,3703	230,9191	✓				
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
68	Dwi Indriyani Sari	115,1029	22,16275	14,75424	242,5277	235,1812			✓		
69	Aulira Putri Maharani	111,9579	18,80326	11,85591	242,6243	235,3446			✓		
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
80	Annisa Alsadila	114,9535	19,00822	15,38871	244,6647	237,6958			✓		
81	Arjun Sandi Wiranata	111,9246	12,04678	14,42654	245,0951	238,2753		✓			
82	Yos Vinoza	115,0788	12,75245	18,40177	245,9586	239,2554		✓			
83	Nabila Maghfirah	112,6166	18,31666	13,91043	244,0379	237,0643			✓		
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
95	Putri Syakina	112,0349	16,3344	17,67237	245,4299	238,5913		✓			
96	Taritiki Amanah W	112,9773	25,59053	17,11359	244,0715	236,9174			✓		
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
105	Arum Sah Putri	111,4966	11,53256	17,33494	247,876	241,3193		✓			
106	Marshella Monica	118,3675	12,82088	30,59207	255,2957	249,5353		✓			
107	Tri Widodo	112,8381	14,59238	21,43449	247,9968	241,5853		✓			
108	Muhammad Syaifudin	113,492	13,4327	15,42522	246,6484	239,9895		✓			
109	Bella Yurice	109,3901	18,00521	16,13033	243,6928	236,6964			✓		
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
122	Pandhu Pramudya	250,8583	247,3962	239,9671	12,66431	17,36663				✓	
123	Rina Delviansta	247,6401	244,1339	236,2859	14,92128	10,90379					✓
157	Yogi Kurnia	257,8172	253,6657	248,0402	16,04421	28,38099				✓	
158	Dandi Cahyo	246,5472	242,1529	235,2718	23,06872	9,22749 6					✓
159	Seli Karyana Aryanti	250,7973	246,4573	240,1892	14,75061	12,86008					✓
160	Carolina Zaitri	254,6473	250,3746	244,4021	12,96488	19,89195				✓	
161	Dena Tasya Vira	250,8922	246,5353	240,3644	14,74277	14,07441					✓
162	Resi Mardila	248,2577	243,919	236,9639	26,02235	10,98075					✓
163	Hanif El-hadi	251,2188	246,86	240,9687	11,70598	19,2395				✓	
164	Maykol Ruanda	245,0239	240,5552	233,7499	26,1403 1	10,90151					✓
165	Risci Ayu Oktaviani	251,6735	247,4186	241,112	16,61552	15,68674					✓
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
173	Benny	248,493	244,161	237,6461	22,11016	16,51044					✓
174	Ari Hidayat	255,9568	251,7365	246,1377	13,92921	26,62679				✓	
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
181	Miranda	251,4726	247,1572	240,5284	18,39278	12,84687					✓
182	Jodi	243,8783	239,4563	233,7017	35,5183	15,67592					✓
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
203	Citia Kartika	250,6998	246,3972	239,984	18,1753	14,35085					✓
204	Cindy Artika Hutabarat	246,8506	242,4563	235,6277	26,06044	10,98439					✓
205	Cindy Amanda Putri	247,9553	243,5722	236,418	25,73578	9,69175 2					✓
206	Dina Utami	251,7456	248,1521	240,6739	12,14312	17,04798				✓	
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
210	Ayu Dewi Retno	253,7217	250,2472	243,0526	7,366171	20,11946				✓	
JUMLAH							148	182	180	175	155
							62	28	30	35	55

Dari tabel 9 didapatkan keanggotaan siswa/siswi SMA Negeri 2 Kota Jambi (iterasi 8). Tidak terjadi perubahan lagi pada setiap *cluster* dan nilai rasio sekarang (0,033508978) sudah tidak lagi lebih besar dari rasio sebelumnya (0,033508978) oleh karena itu algoritma akan dihentikan. Kesimpulan perhitungan penentuan jurusan ke perguruan tinggi untuk siswa/siswi SMA Negeri 2 Kota Jambi dengan cara manual dan hasil rekomendasi jurusannya dikelompokan menjadi 5 cluster diantaranya adalah sebagai berikut :

Cluster 1 siswa/siswi SMA Negeri 2 Kota Jambi direkomendasikan masuk Bidang Kesehatan/kedokteran dalam perhitungan

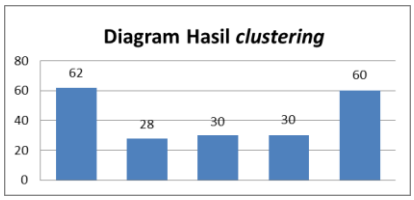
di atas, ada 62 siswa/siswi yang masuk dalam *cluster* 1.

Cluster 2 siswa/siswi siswa/siswi SMA Negeri 2 Kota Jambi direkomendasikan masuk Bidang Agama dalam perhitungan di atas, 28 siswa/siswi yang masuk dalam *cluster* 2.

Cluster 3 siswa/siswi SMA Negeri 2 Kota Jambi direkomendasikan masuk Bidang Teknik dalam perhitungan di atas, ada 30 siswa/siswi yang masuk dalam *cluster* 3.

Cluster 4 siswa/siswi SMA Negeri 2 Kota Jambi direkomendasikan masuk Bidang Pendidikan dan Bidang Seni dalam perhitungan di atas, ada 25 siswa/siswi yang masuk *cluster* 4.

Cluster 5 siswa/siswi SMA Negeri 2 Kota Jambi direkomendasikan masuk Bidang Olahraga dalam perhitungan di atas, ada 55 siswa/siswi yang masuk *cluster* 5.



Gambar 3. grafik hasil *clustering*

Berdasarkan hasil dari hasil penelitian yang telah dilakukan pada SMA Negeri 2 Kota Jambi, maka selain menggunakan perhitungan manual dengan bantuan *software microsoft excel* juga dilakukan perhitungan dengan software data mining menggunakan *tools WEKA* dan *RapidMiner, sebagai* perbandingan hasil yang didapat perhitungan ini adalah sebagai berikut :

Tabel 10. Perbandingan Hasil

Perbandingan	Perhitungan Manual					Perhitungan *RapidMiner*					Perhitungan *WEKA*				
Jumlah iterasi	8					-					10				
Jumlah *Cluster*	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
	62	28	30	35	55	35	55	58	35	27	30	49	41	32	58
Jumlah *Centroid*	5					5					5				
Jumlah Nilai Rasio	30 %	13 %	14 %	17 %	26 %	16,67%	26,19%	27,62%	16,67%	12,86%	14 %	23 %	20 %	15 %	28 %

5. KESIMPULAN

1. Metode yang digunakan dalam penelitian ini adalah *K-Means Clustering* dari perhitungan manual yang telah dilakukan, maka direkomendasikan penulis 5 *cluster*, yang mana untuk *cluster* 1 siswa/siswi SMA Negeri 2 Kota Jambi direkomendasikan masuk Bidang Kesehatan/kedokteran, c*luster* 2 siswa/siswi siswa/siswi SMA Negeri 2 Kota Jambi direkomendasikan masuk Bidang Agama, c*luster* 3 siswa/siswi SMA Negeri 2 Kota Jambi direkomendasikan masuk Bidang Teknik, *cluster* 4 siswa/siswi SMA Negeri 2 Kota Jambi direkomendasikan masuk Bidang Pendidikan dan Bidang Seni, c*luster* 5 siswa/siswi SMA Negeri 2 Kota Jambi direkomendasikan masuk Bidang Olahraga.
2. Pada perhitungan manual terdapat jumlah iterasi sebanyak 8 kali iterasi. Jumlah *cluster* pada perhitungan manual adalah, C1 terdapat 62 data, C2 terdapat 28 data, C3 terdapat 30 data, C4 terdapat 35 data, C5 terdapat 55 data, jumlah *Centroid* pada perhitungan manual adalah 5, Jumlah nilai rasio pada perhitungan manual adalah, C1 terdapat 30%, C2 terdapat 13%, C3 terdapat 14%, C4 terdapat 17%, C5 terdapat 26%.
3. Pada perhitungan *RapidMiner* tidak ditampilkan berapa jumlah iterasi. Jumlah *cluster* pada perhitungan *RapidMiner* adalah, C1 terdapat 35 data, C2 terdapat 55 data, C3 terdapat 58 data, C4 terdapat 35 data, C5 terdapat 27 data, jumlah *centroid* pada perhitungan *RapidMiner* adalah 5, jumlah nilai rasio pada perhitungan *RapidMiner* adalah, C1 terdapat 16,67%, C2 terdapat 26,19%, C3 terdapat 27,62%, C4 terdapat 16,67%, C5 terdapat 12,86%.
4. Pada perhitungan *WEKA* terdapat jumlah iterasi sebanyak 10 kali.

Jumlah *cluster* pada perhitungan *WEKA* adalah, C1 terdapat 30 data, C2 terdapat 49 data, C3 terdapat 41 data, C4 terdapat 32 data, C5 terdapat 58 data, dan jumlah *centroid* pada perhitungan *WEKA* adalah 5, jumlah nilai rasio pada perhitungan *WEKA* adalah, C1 terdapat 14%, C2 terdapat 23%, C3 terdapat 20%, C4 terdapat 15%, C5 terdapat 28%.

DAFTAR PUSTAKA

Florin Gorunescu. (2011). Data Mining Concepts, Models and Techniques. In *The British Journal of Psychiatry* (Vol. 111). https://doi.org/10.1192/bjp.111.479.1009-a

Jiawei Han, Micheline Kamber, J. P. (2011). *Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems)*. Retrieved from http://www.amazon.co.uk/Data-Mining-Concepts-Techniques-Management/dp/0123814790

Jusia, P. A. (2016). Face Recognition Menggunakan Metode Algoritma Viola Jones Dalam Penerapan Computer Vision. *Jurnal Ilmiah Media Processor*, *11*(1), 663–675.

Jusia, P. A. (2017). Decision Support System for Supplier Selection using Analytical Hierarchy Process (AHP) Method. *Scientific Journal of Informatics*, *4*(2), 1–6.

Jusia, P. A. (2018). Analisis komparasi pemodelan algoritma decision tree menggunakan metode particle swarm optimization dan metode adaboost untuk prediksi awal penyakit jantung. *Seminar Nasional Sistem Informasi 2018*, 1048–1056.

Larose, D. T., & Larose, C. D. (2014). *DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining Second Edition Wiley Series on Methods and Applications in Data Mining*.

Quinlan, J. R. (1999). Induction of Decision Trees J.R. *Research and Development in Expert Systems XV*, *1*(Chapter 2), 15–26. https://doi.org/10.1023/A:1022643204877