

BAB II LANDASAN TEORI

2.1. Tinjauan Studi

Tinjauan studi berguna bagi peneliti untuk dijadikan pedoman dan pegangan penelitian yang selanjutnya akan dibuat, nantinya dengan adanya penelitian sebelumnya tersebut memudahkan peneliti sesuai dengan topik pembahasan. Penelitian sebelumnya juga dapat dijadikan perbandingan penelitian sehingga menghasilkan penelitian baru yang lebih bermanfaat. Dalam melaksanakan penelitian ini, diambil beberapa referensi sebagai dasar pelaksanaan penelitian diantaranya :

Menurut Nurzahputra Aldi, Muslim Much Aziz dan Khusniati Miranita, dalam penelitiannya dijelaskan mengenai bagaimana menerapkan Algoritma K-Means dengan berdasarkan Indek Kepuasan Mahasiswa, Data diperoleh dari mahasiswa melalui angket atau kuisioner sesuai aspek Responsiveness, Reliability, Empathy, dan Assurance. Dalam melakukan *clustering*, data yang diperoleh diolah terlebih dahulu dengan cara menghitung berdasarkan bobot yang telah ditentukan. Data akan dihitung terlebih dulu sampai didapat data yang siap untuk dicluster. Penilaian diberikan kepada 146 responden yakni mahasiswa jurusan Ilmu Pengetahuan Alam kepada 12 dosen. Data kemudian diolah menggunakan RapidMinner untuk ditentukan nilai centroid dalam cluster baik dan cluster kurang baik dengan Algoritma K-Means [3].

Menurut Nur Fauziah, Zarlis Muhammad, dan Nasution Benny Benyamin, dalam penelitiannya menjelaskan bagaimana menerapkan algoritma K-Means dalam penentuan jurusan pada siswa baru Sekolah Menengah Kejuruan dengan menggunakan data nilai tes masuk sekolah, penghasilan dan tanggungan anak orang tua. Berdasarkan hasil *cluster* dengan menerapkan beberapa kriteria dari calon siswa menggunakan K-Means dapat diambil pengelompokan dengan rata – rata jurusan yang diambil adalah rekayasa perangkat lunak dan sedikit jumlah siswa yang tidak lulus. Bahkan ada beberapa jurusan yang tidak dibuka dikarenakan kriteria – kriteria siswa tidak dapat lulus dalam jurusan tersebut [2].

Menurut Mardalius, dalam penelitiannya membahas tentang bagaimana implementasi algoritma K-Means *clustering* untuk menentukan kelas kelompok bimbingan belajar tambahan (studi kasus : siswa SMA Negeri 1 Ranah Pesisir), dengan menggunakan data nilai Ujian Nasional SMA jurusan IPA yang meliputi data nilai Kimia, Bahasa Indonesia, Biologi, Matematika, Bahasa Inggris, dan Fisika. Data yang digunakan sejumlah 166 record. Data tersebut akan dikluster menjadi tiga kelompok yaitu kemampuan siswa kurang pintar, siswa sedang, dan siswa pintar. Sehingga dapat diketahui siswa mana yang akan diberi belajar tambahan agar dapat mencapai nilai standar kelulusan Ujian Nasional.

Berdasarkan penelitian tersebut akan dilakukan penelitian dengan menggunakan algoritma yang sama namun dengan format data yang berbeda, sehingga dapat diketahui hasil clustering untuk menentukan siswa bermasalah berdasarkan data Anecdotal Record pada SMK Walisongo Pecangaan. Metode ini dipilih karena mampu mengelompokan data yang memiliki kemiripan karakteristik antara satu data dengan data lain, relatif cepat dan mudah beradaptasi.

2.2. Tinjauan Pustaka

2.2.1. Data Mining

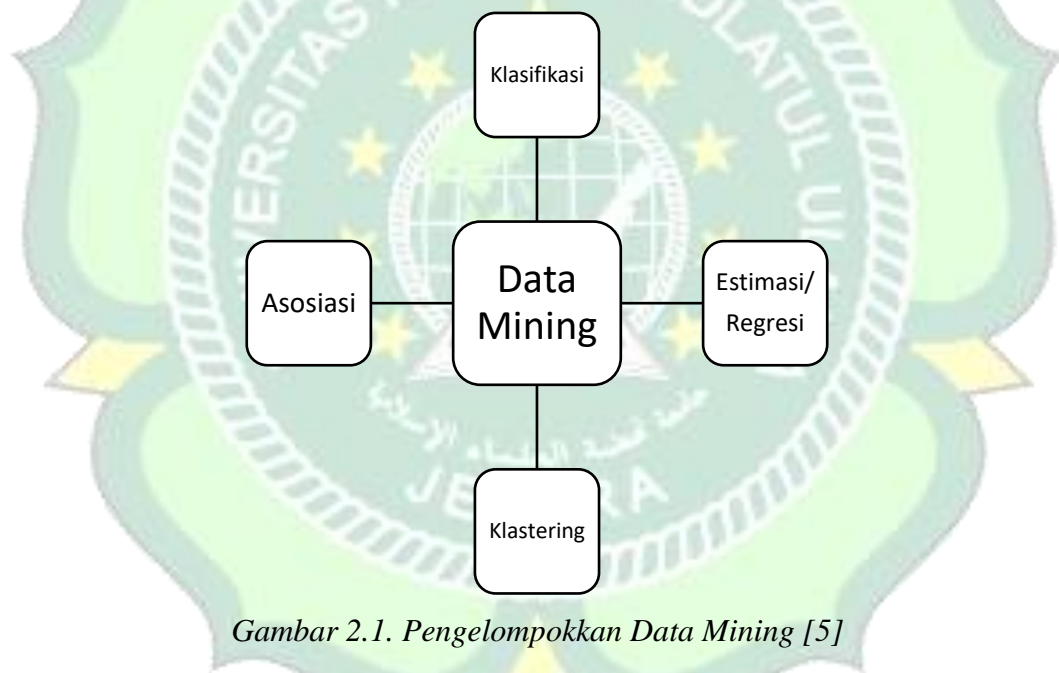
Istilah data mining sudah dikenal sejak tahun 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang penting dalam berbagai bidang, mulai dari bidang bisnis, akademik hingga medis. Munculnya data mining didasarkan pada jumlah datanya yang tersimpan dalam basis data semakin besar. Dalam beberapa teori-teori dan literatur pada data mining sudah ada sejak lama diantaranya *Naïve-Bayes* dan *Nearest Neighbour*, *K-Means Clustering*, aturan asosiasi, *text mining*, dan Pohon Keputusan [2].

Menurut Widodo (2013) Data Mining adalah analisa terhadap data untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut.

Menurut Tan dalam Prasetyo (2012:2) Data Mining adalah suatu proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar.

Berdasarkan dari beberapa pendapat diatas peneliti dapat menarik kesimpulan bahwa Data Mining merupakan suatu proses atau cara untuk mencari pengetahuan yang sebelumnya belum diketahui dengan cara manual dari suatu basis data yang besar dengan aturan tertentu.

Data mining juga biasa disebut dengan *knowledge-discovery in database* (KDD) atau *pattern recognition*. Istilah KDD atau disebut penemuan pengetahuan data karena tujuan utama dari data mining adalah untuk memanfaatkan data dalam basis data dengan mengolahnya sehingga menghasilkan sebuah informasi baru yang berguna. Sedangkan istilah *pattern recognition* atau disebut pengenalan pola yang mempunyai tujuan sebagai pengetahuan yang akan dicari dari dalam sebuah bongkahan data atau dataset yang sedang dihadapi.



Gambar 2.1. Pengelompokan Data Mining [5]

Menurut kusrini (2009) data mining terbagi menjadi beberapa kelompok sebagai berikut :

1) Klastering

Mengelompokkan objek ke dalam beberapa kelompok berdasarkan kemiripan antar objek, di mana dalam satu klaster harus berisi objek yang saling mirip dan antar klaster objek saling tidak mirip. Klaster ini tidak memerlukan data pelatihan yang telah diberi label.

2) Klasifikasi

Melakukan pengelompokan objek berdasarkan kelompok yang sudah ada. Berbeda dengan Klustering, klasifikasi ini memerlukan data pelatihan yang telah diberi label kelompok atau kelas. sebagai contoh kita ingin mengelompokkan data gambar kanker ringan dan akut, maka kita harus menyiapkan misalnya 1000 gambar data pelatihan (data training) dengan label kanker ringan dan 1000 gambar dengan label kanker akut. prediksi pengelompokan dilakukan dengan membangun model terlebih dahulu melalui proses pelatihan menggunakan data yang sudah kita siapkan. Setelah model terbentuk dari proses pelatihan, data baru bisa dikelompokkan menggunakan model tersebut.

3) Estimasi/Regresi

Estimasi pada dasarnya mirip klasifikasi, yakni memerlukan data pelatihan yang sudah diberi label. bedanya, output klasifikasi adalah nilai diskrit, sedangkan output dari estimasi adalah nilai kontinu. titik estimasi ini mencari model hubungan antar atribut prediktor dan atribut dependent, dimana atribut dependennya juga berupa nilai kontinyu. contoh regresi adalah memprediksi nilai kurs Rupiah terhadap Dollar.

4) Asosiasi

Melakukan asosiasi antar objek dalam suatu Set data, biasanya data transaksional. asosiasi dilakukan dengan menghitung Berapa kali dalam suatu Set data suatu transaksi yang mengandung 2 item atau lebih yang berhubungan. sering ada yang menyebut market basket analysis.

2.2.2. Algoritma K-Means

Algoritma K-Means merupakan algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah K Cluster yang sudah ditetapkan di awal. Algoritma K-Means sederhana untuk diimplementasikan dan di jalankan, relative cepat mudah beradaptasi, umum penggunaannya dalam praktek secara historis, K-Means menjadi salah satu algoritma yang paling penting dalam bidang data mining [6].

Algoritma K-Means *Clustering* termasuk dalam salah satu metode data *nonhierarchical clustering* yang dapat mengelompokkan data ke dalam beberapa

cluster berdasarkan kemiripan dari data tersebut. Algoritma K-Means merupakan algoritma teknik *cluster* yang berulang-ulang. Algoritma ini diawali dengan penentuan secara acak K, yang merupakan banyaknya *cluster* yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut dijadikan pusat *cluster* atau biasa disebut dengan *mean* / centroid. Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus yang sudah disediakan hingga ditemukan jarak yang paling dekat dari tiap data terhadap centroid. Klasifikasi dari setiap data berdasarkan kedekatannya dengan centroid. Langkah-langkah tersebut sampai nilai centroid tidak berubah dari sebelumnya [7].

Algoritma k-means adalah algoritma yang mempartisi data ke dalam *cluster* – *cluster* sehingga data yang mempunyai kemiripan berada pada satu *cluster* yang sama sedangkan data yang memiliki ketidaksamaan berada pada cluster yang lain. Langkah-langkah melakukan *clustering* dengan metode K-Means adalah sebagai berikut:



Gambar 2.2. Flowchart Algoritma K-Means Clustering [6]

1. Menentukan K sebagai jumlah kluster yang ingin di bentuk.
2. Membangkitkan nilai acak untuk pusat *cluster* awal (*centroid*) sebanyak Kluster yang telah ditentukan.
3. Menghitung jarak setiap data input terhadap masing – masing centroid menggunakan rumus jarak *Euclidean* (*Euclidean Distance*)

sampai ditemukan jarak yang paling terdekat dari setiap data dengan centroid. Berikut adalah persamaan *Euclidian Distance* :

$$d(c_j, x_i) = \sqrt{\sum(x_i - c_j)^2} \dots\dots\dots (2.1)$$

Dimana :

d : titik dokumen

x_i : data kriteria,

c_j : centroid pada *cluster* ke-j

4. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid (jarak terdekat).
5. Setelah didapat hasil perhitungan data dengan centroid awal setelah itu dicari nilai rata-rata tiap anggota cluster yang nantinya akan digunakan sebagai penentuan centroid baru.
6. Melakukan pengulangan dari langkah 2 sampai 5, sampai anggota tiap cluster tidak ada yang berubah.

Jika langkah 6 telah terpenuhi, maka nilai pusat *cluster* (c_j) pada literasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data [8].

Berikut adalah contoh perhitungan dengan menggunakan metode clustering Algoritma K-Means :

Tabel 2.1. Contoh Dataset

NO	NIS	NAMA SISWA	Pelanggaran					
			B01	B02	B03	B04	B05	B06
1	1160	Fiki Zulfikar	2	2	1	0	0	0
2	1822	Alvin Latiful Qohar	2	2	0	0	0	0
3	1875	Achmad Robiansyah	1	1	0	0	0	0
4	1876	Adinda Ayu Wardani	0	0	0	0	0	0
5	1878	Ainul Falaq	1	0	0	0	0	0

Sebelum dilakukan Clustering, data yang diperoleh akan di olah terlebih dahulu sesuai bobot atau poin yang telah ditentukan. Data tersebut akan dihitung sehingga didapatkan data yang siap dicluster.

Tabel 2.2. Data Pembobotan

ID	Nama	Poin
B01	Datang terlambat < 10 menit	1
B02	Datang terlambat < 30 menit	2
B03	Datang terlambat > 30 Menit	3
B04	Tidak mengikuti pelajaran tanpa ijin	4
B05	Tidak mengerjakan tugas Pekerjaan Rumah/tugas lain	3
B06	Tidak mengikuti kegiatan ekstra kurikuler	2

Berikut tabel proses pembobotan berdasarkan poin yang telah ditentukan, disajikan pada tabel 2.3 berikut:

Tabel 2.3. Proses pembobotan

ID	(Jumlah x Bobot)	Hasil
B01	2 x 1	2
B02	2 x 2	4
B03	1 x 3	3
B04	0 x 4	0
B05	0 x 3	0
B06	0 x 2	0
Jumlah		9

Tabel 2.4. Nilai Siswa setelah perhitungan

NIS	Hasil
1660	9
1822	6
1875	3
1876	0
1878	1

Data yang telah diakumulasikan selanjutnya akan dicluster menjadi dua cluster, sebelumnya kita tentukan centroid awal terlebih dahulu. Kita ambil data ke-1 dan ke-2 sebagai perhitungan pertama. Kita akan menggunakan persamaan *Euclidean Distance* untuk memperoleh jarak minimum data terhadap centroid.

Tabel 2.5. Centroid awal

Cluster	x
C1	6
C2	9

Literasi Ke-1

$$S1 \quad d(x_1, c_1) = \sqrt{(9-6)^2} = 3 \quad d(x_1, c_2) = \sqrt{(9-9)^2} = 0$$

$$S2 \quad d(x_2, c_1) = \sqrt{(6-6)^2} = 0 \quad d(x_2, c_2) = \sqrt{(6-9)^2} = 3$$

$$S3 \quad C1 \quad d(x_3, c_1) = \sqrt{(3-6)^2} = 3 \quad C2 \quad d(x_3, c_2) = \sqrt{(3-9)^2} = 6$$

$$S4 \quad d(x_4, c_1) = \sqrt{(0-6)^2} = 6 \quad d(x_4, c_2) = \sqrt{(0-9)^2} = 9$$

$$S5 \quad d(x_5, c_1) = \sqrt{(1-6)^2} = 5 \quad d(x_5, c_2) = \sqrt{(1-9)^2} = 8$$

Dari perhitungan data diatas maka diperoleh data nilai siswa S1 dinyatakan masuk C2 dan S2,S3,S4, dan S5 masuk C1, sesuai hasil nilai yang terkecil.

Tabel 2.6. Pengelompokan data pada Literasi ke-1

ID	C1	C2	Cluster
S1	3	0	C2
S2	0	3	C1
S3	3	6	C1
S4	6	9	C1
S5	5	8	C1

Setelah mendapatkan label cluster untuk masing-masing data maka dicari nilai rata-ratanya dengan menjumlahkan seluruh anggota masing-masing cluster dan dibagi jumlah anggotanya.

Tabel 2.7. Centroid baru pada literasi ke-1

Cluster	x
C1	2.5
C2	9

Literasi Ke-2

$$\begin{array}{ll}
 \text{S1} & d(x_1, c_1) = \sqrt{(9 - 2.5)^2} = 6.5 & d(x_1, c_2) = \sqrt{(9 - 9)^2} = 0 \\
 \text{S2} & d(x_2, c_1) = \sqrt{(6 - 2.5)^2} = 3.5 & d(x_2, c_2) = \sqrt{(6 - 9)^2} = 3 \\
 \text{S3} & d(x_3, c_1) = \sqrt{(3 - 2.5)^2} = 0.5 & \text{C2} \quad d(x_3, c_2) = \sqrt{(3 - 9)^2} = 6 \\
 \text{S4} & d(x_4, c_1) = \sqrt{(0 - 2.5)^2} = 2.5 & d(x_4, c_2) = \sqrt{(0 - 9)^2} = 9 \\
 \text{S5} & d(x_5, c_1) = \sqrt{(1 - 2.5)^2} = 1.5 & d(x_5, c_2) = \sqrt{(1 - 9)^2} = 8
 \end{array}$$

Dari perhitungan data diatas maka dapat diketahui bahwa data nilai siswa S1 dan S2 dinyatakan masuk C2 dan S3,S4, dan S5 masuk C1, sesuai hasil nilai yang terkecil.

Tabel 2.8. Pengelompokan data pada Literasi ke-2

ID	C1	C2	Cluster
S1	6.5	0	C2
S2	3.5	3	C2
S3	0.5	6	C1
S4	3.5	9	C1
S5	1.8	8	C1

Setelah mendapatkan label cluster untuk masing-masing data maka dicari nilai rata-ratanya dengan menjumlahkan seluruh anggota masing-masing cluster dan dibagi jumlah anggotanya.

Tabel 2.9. Centroid baru pada literasi ke-2

Cluster	x
C1	1.33
C2	7.5

Literasi Ke-3

$$\begin{array}{ll}
 \text{S1} & d(x_1, c_1) = \sqrt{(9 - 1.33)^2} = 8.8 & d(x_1, c_2) = \sqrt{(9 - 7.5)^2} = 1.5 \\
 \text{S2} & d(x_2, c_1) = \sqrt{(6 - 1.33)^2} = 5.8 & d(x_2, c_2) = \sqrt{(6 - 7.5)^2} = 1.5 \\
 \text{S3} & d(x_3, c_1) = \sqrt{(3 - 1.33)^2} = 2.8 & \text{C2} \quad d(x_3, c_2) = \sqrt{(3 - 7.5)^2} = 4.5 \\
 \text{S4} & d(x_4, c_1) = \sqrt{(0 - 1.33)^2} = 0.2 & d(x_4, c_2) = \sqrt{(0 - 7.5)^2} = 7.5 \\
 \text{S5} & d(x_5, c_1) = \sqrt{(1 - 1.33)^2} = 0.8 & d(x_5, c_2) = \sqrt{(1 - 7.5)^2} = 6.5
 \end{array}$$

Dari perhitungan data diatas maka dapat disimpulkan bahwa data nilai siswa S1 dan S2 dinyatakan masuk C2 dan S3,S4, dan S5 masuk C1, sesuai hasil nilai yang terkecil.

Tabel 2.10. Pengelompokan data pada Literasi ke-3

ID	C1	C2	Cluster
S1	7.66	1.5	C2
S2	4.66	1.5	C2
S3	1.66	4.5	C1
S4	1.33	7.5	C1
S5	0.33	6.5	C1

Setelah mendapatkan label cluster untuk masing-masing data maka dicari nilai rata-ratanya dengan menjumlahkan seluruh anggota masing-masing cluster dan dibagi jumlah anggotanya.

Tabel 2.11. Centroid baru pada literasi ke-3

Cluster	x
C1	1.33
C2	7.5

Karena centroid tidak mengalami perubahan (sama dengan centroid sebelumnya) maka proses clustering selesai.

2.2.3. Davies-Buildin Index

Matrik Davies-Buildin Index (DBI) diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 yang digunakan untuk mengevaluasi cluster. Validitas yang dilakukan adalah seberapa baik clustering yang sudah dilakukan dengan cara menghitung kuantitas dan fitur turunan dari set data [6].

Sum of Square within cluster (SSW) sebagai metrik kohesi dalam sebuah cluster ke-*i* diformulasikan oleh persamaan (2.2)

$$SSW = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \dots\dots\dots (2.2)$$

m_i adalah jumlah data yang berada dalam cluster ke- i , sedangkan c_i adalah centroid ke- i .

sementara metrik untuk separasi antar dua cluster, misalnya i dan j , digunakan formula *sum of square between cluster* (SSB) dengan mengukur jarak antar centroid c_i dan c_j seperti pada persamaan berikut :

$$SBB_{i,j} = d(c_i, c_j) \dots\dots\dots (2.3)$$

Didefinisikan R_{ij} adalah rasio seberapa baik nilai perbandingan antar cluster ke- i dan cluster ke- j . Nilainya didapatkan dari komponen kohesi dan separasi. Cluster yang baik adalah yang mempunyai kohesi yang sekecil mungkin dan sparasi yang sebesar mungkin. R_{ij} diformulasikan oleh persamaan berikut :

$$R_{i,j} = \frac{SSW_i SSW_j}{SSB_{i,j}} \dots\dots\dots (2.4)$$

Sifat-sifat yang dimiliki R_{ij} sebagai berikut :

- $R_{ij} \geq 0$
- $R_{i,j} = R_{j,i}$
- Jika $SSW_j \geq SSW_r$ dan $SSB_{i,j} = SSB_{i,r}$ maka $R_{i,j} > R_{i,r}$
- Jika $SSW_j = SSW_r$ dan $SSB_{i,j} \leq SSB_{i,r}$ maka $R_{i,j} > R_{i,r}$

Nilai Davies-Bouldin Index (DBi) didapatkan dari persamaan berikut :

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{i,j}) \dots\dots\dots (2.5)$$

K adalah jumlah cluster yang digunakan.

Dari syarat-syarat perhitungan yang didefinisikan di atas, dapat diamati bahwa semakin kecil nilai SSW maka hasil clustering yang didapat juga akan baik. Secara esensial, DBI menginginkan nilai dengan sekecil (non-negatif ≥ 0) mungkin untuk menentukan baik atau tidaknya cluster yang didapat. Indeks tersebut diperoleh dari rata-rata semua indeks cluster, dan nilai yang didapat dapat digunakan sebagai pendukung keputusan untuk penentuan jumlah cluster yang cocok digunakan.

Hasil Clustering seperti pada contoh sebelumnya pada algoritma K-Means akan dilakukan validasi dengan Davies-Bouldin Index. Data hasil clustering disajikan pada Tabel 2.12, sedangkan centroid setiap cluster disajikan pada table 2.13 parameter jarak yang digunakan adalah *Euclidian Distance*.

Tabel 2.12. Hasil Clustering pada contoh K-means

ID	x	Cluster
S1	9	C2
S2	6	C2
S3	3	C1
S4	0	C1
S5	1	C1

Tabel 2.13. Centroid yang diperoleh dari contoh Clustering

Cluster	x
C1	1.33
C2	7.5

Tabel 2.14 menyajikan proses perhitungan SSW. Pada tabel tersebut dikelompokkan berdasarkan cluster-nya, kemudian dihitung jarak setiap centroid masing-masing dan dihitung rata-ratanya untuk menjadi SSW. Contoh perhitungan jarak data dalam cluster 1 ke centroid-nya sebagai berikut :

$$d(x_3, c_1) = \sqrt{(3 - 1.33)^2} = 1.67$$

$$d(x_4, c_1) = \sqrt{(0 - 1.33)^2} = 1.33$$

$$d(x_5, c_1) = \sqrt{(1 - 1.33)^2} = 0.33$$

$$d(x_1, c_2) = \sqrt{(9 - 7.5)^2} = 1.5$$

$$d(x_2, c_2) = \sqrt{(6 - 7.5)^2} = 1.5$$

Sementara SSW untuk cluster 1 didapat sebagai berikut :

$$\begin{aligned} SSW_1 &= \frac{1}{m} (d(x_3, c_1) + d(x_4, c_1) + d(x_5, c_1)) \\ &= \frac{1}{3} (1.67 + 1.33 + 0.33) = 1.11 \end{aligned}$$

Nilai SSW untuk cluster 2 dihitung dengan cara yang sama seperti di atas

$$SSW_2 = \frac{1}{m} (d(x_1, c_2) + d(x_2, c_2)) = \frac{1}{2} (1.5 + 1.5) = 1.5$$

Tabel 2.14. Perhitungan SSW

ID	x	cluster	centroid	Jarak	SSW
S1	9	C2	7.5	1.5	1.5
S2	6			1.5	
S3	3	C1	1.33	1.67	1.11
S4	0			1.33	
S5	1			0.33	

SSB didapatkan dengan menghitung jarak antar centroid, hasilnya disajikan seperti pada Tabel 2.15. Contoh perhitungan SSB pasangan diantara dua cluster tersebut sebagai berikut :

$$SSB_{1,2} = d(c_1 c_2) = \sqrt{d(x_{c1} - x_{c2})^2}$$

$$= \sqrt{(1.33 - 7.5)^2} = 6.17$$

Tabel 2.15. Perhitungan SSB

		Data ke-i	
		SSB	
Data ke-i	1	0	6.17
	2	6.17	0

Berikut nilai R yang didapatkan dari persamaan (2.4) :

$$R_{1,2} = \frac{SSW_1 + SSW_2}{SSB_{1,2}} = \frac{1.5 + 1.11}{6.17} = 0.42$$

Hasil DBI yang didapatkan dari persamaan (2.5). Hasilnya ditampilkan seperti dibawah ini :

Tabel 2.16. Perhitungan R dan DBI

		Data ke-i		R Max	DBI
		R			
Data ke-i	1	0	0.42	0.42	0.42
	2	0.42	0	0.42	

Dari Tabel 2.16, terlihat bahwa nilai DBI yang didapat adalah 0.42.

2.2.4. Anecdotal Record

Catatan menurut Kamus Besar Bahasa Indonesia adalah hasil mencatat, sedangkan anekdot adalah catatan tentang kejadian yang berkaitan dengan masalah yang sedang menjadi pusat perhatian pengamat, terutama catatan tentang tingkah laku individu yang bersifat khas [8].

Anecdotal Record atau catatan anekdot adalah penggambaran secara tertulis dari perilaku anak [9]. Selain itu catatan anekdot juga dapat diartikan dengan suatu deskripsi atau catatan rekaman tentang peristiwa-peristiwa yang terjadi dalam situasi natural atau alamiah. Lazimnya pencatatan peristiwa ini difokuskan pada peserta didik yang sedang menjadi perhatian guru, sehingga himpunan dari catatan-catatan anekdot semacam ini akan memberikan deskripsi atau gambaran tentang pola tingkah laku peserta didik yang bersangkutan.

Anecdotal Record ditulis dengan singkat. Anecdotal Record menjelaskan sesuatu yang terjadi secara faktual yaitu sesuai dengan apa yang dilihat dan didengar, dengan cara yang obyektif atau tidak berprasangka, tidak menduga-duga, menceritakan kapan, di mana, dan bagaimana terjadinya peristiwa itu, serta apa yang dilakukan dan dikatakan anak [1].

Perilaku seseorang pada umumnya mendeskripsikan kecenderungan atau kebiasaan seseorang dalam suatu hal. Misalnya, seseorang yang biasanya minum kopi dapat diketahui sebagai kecenderungannya yang suka pada kopi. Oleh karena itu, guru dapat melakukan pemantauan terhadap peserta didik yang dibinanya. Hasil pemantauan tersebut dapat digunakan sebagai umpan balik atau acuan dalam pembinaan. Observasi perilaku disekolah bisa dilakukan dengan cara menggunakan buku catatan khusus yang berkaitan dengan kejadian-kejadian yang diperbuat oleh peserta didik selama disekolah. Catatan dalam buku tersebut, selain berguna untuk menilai dan merekam perilaku peserta didik juga bisa untuk menilai sikap dari peserta didik serta menjadi bahan dalam penilaian perkembangan secara menyeluruh.

Dibawah ini adalah contoh dari penulisan Anecdotal Record :

Kelas : VIII

Nama murid : Ahmadi

Catatan ke : 3

Tanggal : 28 Agustus 2019

Waktu/pukul : 09.00

Tempat : Di dalam kelas

Ahmadi terlambat masuk kelas 20 menit. Dia mengutarakan alasannya terlambat, bahwa dia harus menunggu lama giliran mendapatkan pesanan makanan untuk sarapannya di kantin sekolah. Sesudah duduk, dia mengeluarkan alat tulis dan bukunya dan mengerjakan tugas yang diberikan oleh guru. Sesudah menyelesaikan 6 dari 10 soal matematika yang ditugaskan oleh guru, dia menaruh kepalanya di meja. Setelah kurun waktu beberapa menit, dia tertidur.

2.2.5. RapidMiner

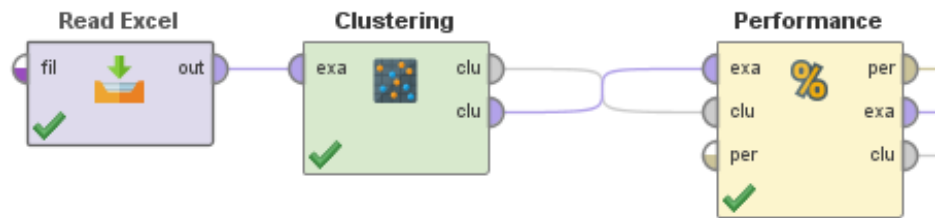
RapidMiner adalah sebuah *software* yang bersifat *Open Source* yang dapat melakukan analisis terhadap sebuah data mining. RapidMiner sebagai salah satu *software* untuk pengolah data mining yang menyediakan berbagai tool untuk membuat sebuah *decision tree*.

RapidMiner menggunakan teknik prediksi dan deskriptif dalam memberikan pengetahuan kepada pengguna sehingga aplikasi ini dapat membuat sebuah keputusan yang baik. RapidMiner memiliki operator data mining sekitar 500 operator. RapidMiner dibuat dengan menggunakan bahasa pemrograman java sehingga dapat berjalan di berbagai *Operating System*.

RapidMiner sebelumnya diberi nama YALE (*Yet Another Learning Environment*), di mana versi pertamanya dikembangkan pada tahun 2001 oleh Ingo Mierswa, Simon Fischer, dan RalfKlinkenberg di Artificial Intelligence Unit dari University of Dortmund Germany. RapidMiner didistribusikan di bawah lisensi AGPL (*GNU Affero General Public License*) versi 3. Hingga saat ini telah banyak aplikasi yang dikembangkan menggunakan RapidMiner di lebih dari 40 negara. Walaupun RapidMiner adalah *software open source* sebagai pengolahan data mining, akan tetapi *software* ini tidak perlu diragukan karena sudah terkemuka di dunia [7].

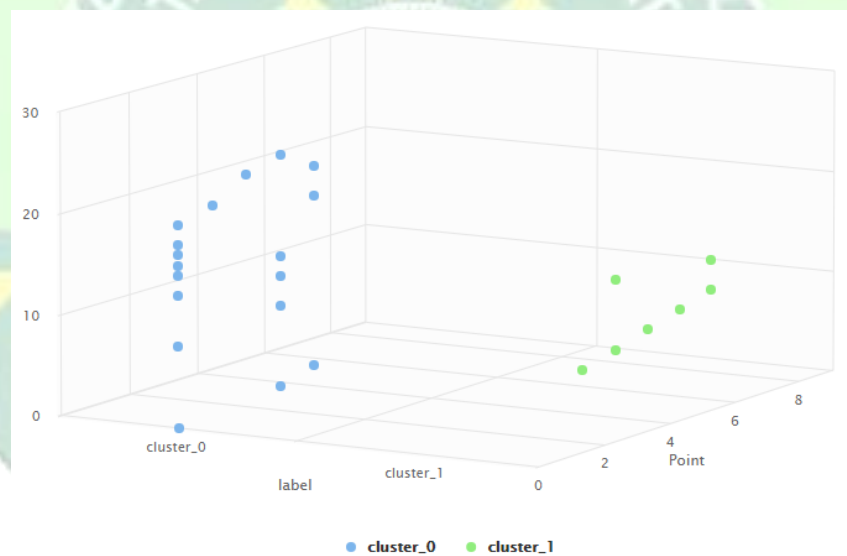
RapidMiner menyediakan tampilan GUI (*Graphic User Interface*) untuk merancang sebuah *pipeline* analitis. GUI ini akan membuat file XML (*Extensible Markup Language*) yang menginterpretasikan proses analitis

keinginan dari pengguna untuk diimplemetasikan ke data. Kemudian file tersebut akan diproses oleh RapidMiner untuk melakukan analis secara otomatis.



Gambar 2.3. Metode K-Means dengan Rapidminer

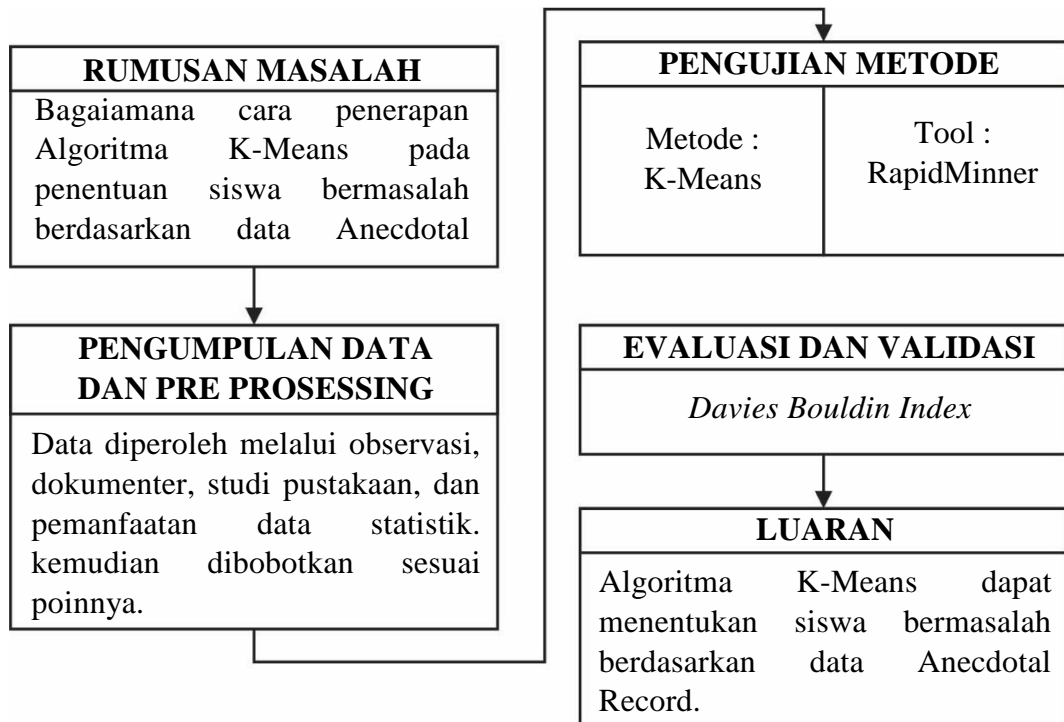
Gambar 2.3 di mulai dari pembacaan dataset yang dimasukkan dengan melalui Microsoft Excel yang sebelumnya telah diklasifikasi variabelnya, Operator Clustering sebagai model operator untuk metode K-Means digunakan untuk perhitungan jarak dari setiap kluster data, operator *Performance* sebagai pengukuran validasi menggunakan metode *Index Davis Bouldin*.



Gambar 2.4. Hasil Clustering

2.3. Kerangka Pemikiran

Pada tahapan ini peneliti membuat suatu kerangka pemikiran secara bertahap. Kerangka pemikiran merupakan pola pikir atau konsep dalam melakukan suatu penelitian.



Gambar 2.5. Kerangka Pemikiran