
**Analisis Data Mining Absensi Siswa SMP Negeri 1 Bilah Barat
Dengan Metode Algoritma K-Means Clustering**

Shondy Raja Ferdinand Sinurat¹, Marnis Nasution², Rahma Muti Ah³

Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu^{1,2,3}

E-mail: rajasinurat516@gmail.com¹, marnisnst@gmail.com²,
rmhutea5@gmail.com³

Corresponding Author: rajasinurat516@gmail.com

Abstract

The presence of students in Junior High School has significant implications in the world of Education. The success of education depends not only on the content of the curriculum and the quality of teaching, but is also strongly linked to the extent to which students are regularly present at school. The problem of student absenteeism in junior high school is an issue that should not be ignored, as it can negatively affect academic achievement, the development of social skills, and the overall educational experience of students. A student's absence from school can be caused by a complex of factors. These factors include personal problems such as lack of motivation, family problems such as domestic conflicts, poor physical or mental health, as well as environmental factors such as school accessibility. A deep understanding of the reasons behind student absenteeism is an important first step to addressing this problem. Student attendance Data that has been accumulated over the past 6 months can be a valuable source of information in analyzing student attendance. However, extracting useful information from large and complex attendance data is a challenging task. This is where Data Mining comes in. Data Mining is an approach that enables the identification of patterns, trends, and valuable information in large and complex data.

Keywords: SMP Negeri 1 Bilah Barat, Analysis, Data Mining, Attendance, Algorithm K-Means Clustering.

I. Pendahuluan

Kehadiran siswa di Sekolah Menengah Pertama (SMP) memiliki implikasi signifikan dalam dunia pendidikan. Keberhasilan pendidikan tidak hanya tergantung pada isi kurikulum dan kualitas pengajaran, tetapi juga sangat terkait dengan sejauh mana siswa hadir secara teratur di sekolah. Masalah absensi siswa di SMP adalah isu yang tidak boleh diabaikan, karena dapat berdampak negatif pada

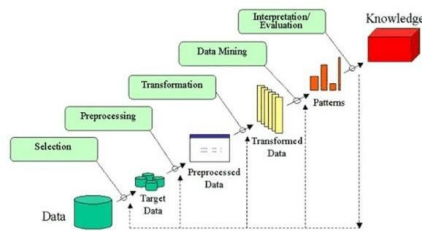
pencapaian akademik, perkembangan keterampilan sosial, dan keseluruhan pengalaman pendidikan siswa. Faktor-faktor ini mencakup masalah pribadi seperti kurangnya motivasi, masalah keluarga seperti konflik rumah tangga, kesehatan fisik atau mental yang buruk, serta faktor lingkungan seperti aksesibilitas sekolah. Pemahaman yang mendalam tentang alasan di balik ketidakhadiran siswa adalah langkah awal yang penting untuk mengatasi

masalah ini. Data absensi siswa yang telah terakumulasi selama 6 bulan terakhir dapat menjadi sumber informasi yang berharga dalam menganalisis kehadiran siswa. Dalam penelitian ini, kami akan fokus pada penggunaan metode Data Mining, khususnya Algoritma K-means Clustering, untuk menganalisis data absensi siswa di SMP Negeri 1 Bilah Barat. Algoritma K-means Clustering telah terbukti berhasil dalam analisis kluster, di mana data dikelompokkan berdasarkan kesamaan mereka. Dalam konteks absensi siswa,

setiap dalam cluster tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam cluster yang lainnya. Sampai sekarang, para ilmuwan masih terus melakukan berbagai usaha untuk melakukan berbagai usaha untuk melakukan perbaikan model cluster dan menghitung jumlah cluster yang optimal sehingga dapat dihasilkan cluster yang paling baik. Ada dua metode clustering yang dikenal, yaitu *hierarchical clustering* dan *partitioning*.

II. Landasan Teori
Knowledge Discovery In Database

Knowledge Discovery In Database (KDD) merupakan nama lain dari Data Mining walaupun sesungguhnya kedua istilah tersebut memiliki konsep yang tidak sama, namun berkaitan satu sama lain, dan salah satu dari tahapan proses keseluruhan *Knowledge Discovery In Database* (KDD) merupakan Data Mining yang menjadi inti dari proses KDD.



Gambar 1. Ilustrasi Proses
Knowledge Discovery In Database

Clustering

Salah satu yang dikenal dalam data mining yaitu clustering. Pengertian clustering keilmuan dalam data mining adalah pengelompokkan sejumlah data objek kedalam cluster (group) sehingga

K-Means

K-Means merupakan salah satu metode data clustering yang non hierarki yang berusaha mempartisi data yang ada kedalam bentuk satu maupun lebih cluster atau kelompok sehingga data yang memiliki karakteristik yang sama di kelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok lainnya.

III. Metode Penelitian

Metode yang dipakai pada penelitian ini adalah K-Means Clustering akan mencari partisi yang maksimum dari data dengan meminimalkan kriteria jumlah kesalahan kuadrat dengan prosedur iterasi yang optimal. Variabel yang digunakan pada penelitian ini adalah data absensi siswa selama 6 bulan terakhir dan data siswa. Tujuan dari metode penelitian ini adalah mengelompokkan siswa di SMP Negeri 1 Bilah Barat berdasarkan absensi, dan mengetahui pola absensi untuk meningkatkan mutu belajar.

IV. Hasil Dan Pembahasan Langkah – Langkah pengelolaan Algoritma K – Means Clustering Data

Data yang digunakan dalam penelitian ini adalah data absensi siswa dan data siswa yang diperoleh dari sekolah tepatnya pada Tata Usaha dan Guru BK sekolah. Data – data tersebut berisi data absensi selama 6 bulan terakhir seperti nama siswa, kelas, absen perbulan. Jumlah data yang digunakan dalam penelitian ini 84 data siswa dan absensi dari kelas VIII-A, VIII-B, VIII-C, dan data absensi dari bulan januari 2023 sampai juni 2023.

Tabel 1. Sampel Data

Nama	Jumlah Hari Absen	Jenis Absensi	Pola Absensi	Usia	Jenis Kelamin	Kelas
Abgan Pratama. F	6	Ijin, Sakit, Tanpa Keterangan	Senin – Sabtu	13	Laki – Laki	8A
Aldo Febriand	6	Ijin, Sakit, Tanpa Keterangan	Senin – Sabtu	13	Laki – Laki	8B
Andin Flower. S	6	Ijin, Sakit, Tanpa Keterangan	Senin – Sabtu	13	Perempuan	8C
Arifin Hanuy P	6	Ijin, Sakit, Tanpa Keterangan	Senin – Sabtu	13	Perempuan	8A
Chelsi Aulia. F. M	6	Ijin, Sakit, Tanpa Keterangan	Senin – Sabtu	13	Perempuan	8B
Claudia Margaret. S	6	Ijin, Sakit, Tanpa Keterangan	Senin – Sabtu	13	Perempuan	8C
Diva Novianti Sianga	6	Ijin, Sakit, Tanpa Keterangan	Senin – Sabtu	13	Perempuan	8A
Fadli Pradana	6	Ijin, Sakit, Tanpa Keterangan	Senin – Sabtu	13	Laki – Laki	8B
Greis Olivia J. M	6	Ijin, Sakit, Tanpa Keterangan	Senin – Sabtu	13	Perempuan	8C
Hanih Siregar	6	Ijin, Sakit, Tanpa Keterangan	Senin – Sabtu	13	Perempuan	8A

Pra-Proses Data

Pra-Proses Data adalah salah satu Langkah penting dalam proses data mining yang berkaitan dengan persiapan dan transformasi data set awal.

Transformasi Data

Transformasi data adalah proses perubahan data, menjadi data yang dapat diolah menggunakan algoritma yang akan dipakai apakah dalam bentuk numerik, kasifikasi dan lain – lain. Agar data diatas dapat diolah dengan

menggunakan algoritma *K-Means*, maka data yang berjenis data nominal seperti nama, jarak rumah, jumlah absensi, dan penghasilan orang tua. di inisialisasikan terlebih dahulu dalam bentuk angka satuan.

Tabel 2. Sampel Pra-Proses data

Nama	Januari	Februari	Maret	April	Mei	Juni
K1	20	18	22	21	20	19
K2	22	23	24	25	22	23
K3	18	19	20	21	22	23
K4	16	17	18	17	16	15
K5	25	24	23	24	25	26
K6	19	20	19	18	19	20
K7	21	20	22	23	24	25
K8	17	16	18	17	16	15
K9	23	22	24	25	24	23
K10	20	21	20	21	22	21

Analisis data dengan Algoritma *K-Means Clustering*

K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada kedalam bentuk satu atau lebih cluster/kelompok. Metode ini mempunyai karakteristik yang berbeda di Clustering kedalam kelompok yang lain.

Selanjutnya pengolahan data untuk mendapatkan informasi pola-pola absensi yang memiliki potensial dalam absensi sekolah. Proses kali ini hanya menggunakan tiga kluster dan atribut data yang digunakan adalah nama, jarak tempuh siswa, jumlah kehadiran, penghasilan orang tua. Berikut pada tabel 3 adalah titik pusat awal kluster.

1. Tentukan Jumlah Kluster k.
2. Menentukan Titik pusat awal dari setiap kluster

Tabel 3. Titik Pusat Awal Kluster

Nama	Januari	Februari	Maret	April	Mei	Juni	
K1	20	18	22	21	20	19	C1
K5	25	24	23	24	25	26	C2

3. Selanjutnya adalah menghitung jarak terdekat, dikarenakan metode *K-Means* mengalokasikan setiap data kedalam suatu kluster dari data

yang memiliki jarak paling dekat dengan titik pusat dari setiap kluster, maka dari itu perlu dihitung jarak setiap data dengan menggunakan rumus Encludian Distance, sebagai contoh akan dihitung jarak data pertama kepusat kluster pertama:

$$D(x,y) = \sum_{i=1}^n (x_i - y_i)^2$$

Data K1

$$\begin{aligned} D(K_1, C_1) &= \sqrt{(20 - 20)^2 + (18 - 18)^2 + (22 - 22)^2 + (21 - 21)^2 + (20 - 20)^2 + (19 - 19)^2} \\ &= \sqrt{0} \\ &= 0 \end{aligned}$$

$$\begin{aligned} D(K_1, C_2) &= \sqrt{(20 - 25)^2 + (18 - 24)^2 + (22 - 23)^2 + (21 - 24)^2 + (20 - 25)^2 + (19 - 26)^2} \\ &= \sqrt{(-5)^2 + (-6)^2 + (-1)^2 + (-3)^2 + (-5)^2 + (-7)^2} \\ &= \sqrt{25 + 36 + 1 + 9 + 25 + 49} \\ &= \sqrt{145} = 12.04 \end{aligned}$$

Data K2

$$\begin{aligned} D(K_1, C_1) &= \sqrt{(22 - 20)^2 + (23 - 18)^2 + (24 - 22)^2 + (25 - 21)^2 + (22 - 20)^2 + (23 - 19)^2} \\ &= \sqrt{(2)^2 + (5)^2 + (2)^2 + (4)^2 + (2)^2 + (4)^2} \\ &= \sqrt{4 + 25 + 4 + 16 + 4 + 16} \\ &= \sqrt{69} = 8.31 \end{aligned}$$

$$\begin{aligned} D(K_1, C_2) &= \sqrt{(22 - 25)^2 + (23 - 24)^2 + (24 - 23)^2 + (25 - 24)^2 + (22 - 25)^2 + (23 - 26)^2} \\ &= \sqrt{(-3)^2 + (-1)^2 + (1)^2 + (1)^2 + (-3)^2 + (-3)^2} \\ &= \sqrt{9 + 1 + 1 + 1 + 9 + 9} \\ &= \sqrt{30} = 5.48 \end{aligned}$$

Data K3

$$\begin{aligned} D(K_3, C_1) &= \sqrt{(18 - 20)^2 + (19 - 18)^2 + (20 - 22)^2 + (21 - 21)^2 + (22 - 20)^2 + (23 - 19)^2} \\ &= \sqrt{(-2)^2 + (1)^2 + (-2)^2 + (0)^2 + (2)^2 + (4)^2} \\ &= \sqrt{4 + 1 + 4 + 0 + 4 + 16} \\ &= \sqrt{29} = 5.39 \end{aligned}$$

$$\begin{aligned} D(K_3, C_2) &= \sqrt{(18 - 25)^2 + (19 - 24)^2 + (20 - 23)^2 + (21 - 24)^2 + (22 - 25)^2 + (23 - 26)^2} \\ &= \sqrt{(-7)^2 + (-5)^2 + (-3)^2 + (-3)^2 + (-3)^2 + (-3)^2} \\ &= \sqrt{49 + 25 + 9 + 9 + 9 + 9} \\ &= \sqrt{110} = 10.49 \end{aligned}$$

Data K4

$$\begin{aligned} D(K_4, C_1) &= \sqrt{(16 - 20)^2 + (17 - 18)^2 + (18 - 22)^2 + (17 - 21)^2 + (16 - 20)^2 + (15 - 19)^2} \\ &= \sqrt{(-4)^2 + (-1)^2 + (-4)^2 + (-4)^2 + (-4)^2 + (-4)^2} \\ &= \sqrt{16 + 1 + 16 + 16 + 16 + 16} \\ &= \sqrt{81} = 9 \end{aligned}$$

$$\begin{aligned} D(K_4, C_2) &= \sqrt{(16 - 25)^2 + (17 - 24)^2 + (18 - 23)^2 + (17 - 24)^2 + (16 - 25)^2 + (15 - 26)^2} \\ &= \sqrt{(-9)^2 + (-7)^2 + (-5)^2 + (-7)^2 + (-9)^2 + (-11)^2} \\ &= \sqrt{81 + 49 + 25 + 49 + 81 + 121} \\ &= \sqrt{406} = 20.14 \end{aligned}$$

Data K5

$$\begin{aligned} D(K_5, C_1) &= \sqrt{(25 - 20)^2 + (24 - 18)^2 + (23 - 22)^2 + (24 - 21)^2 + (25 - 20)^2 + (26 - 19)^2} \\ &= \sqrt{(5)^2 + (6)^2 + (1)^2 + (3)^2 + (5)^2 + (7)^2} \\ &= \sqrt{25 + 36 + 1 + 9 + 25 + 49} \\ &= \sqrt{145} = 12.04 \end{aligned}$$

$$\begin{aligned} D(K_5, C_2) &= \sqrt{(25 - 25)^2 + (24 - 24)^2 + (23 - 23)^2 + (24 - 24)^2 + (25 - 25)^2 + (26 - 26)^2} \\ &= 0 \end{aligned}$$

Data K6

$$\begin{aligned} D(K_6, C_1) &= \sqrt{(19 - 20)^2 + (20 - 18)^2 + (19 - 22)^2 + (18 - 21)^2 + (19 - 20)^2 + (20 - 19)^2} \\ &= \sqrt{(-1)^2 + (2)^2 + (-3)^2 + (-3)^2 + (-1)^2 + (1)^2} \\ &= \sqrt{1 + 4 + 9 + 9 + 1 + 1} \\ &= \sqrt{25} = 5 \end{aligned}$$

$$\begin{aligned} D(K_6, C_2) &= \sqrt{(19 - 25)^2 + (20 - 24)^2 + (19 - 23)^2 + (18 - 24)^2 + (19 - 25)^2 + (20 - 26)^2} \\ &= \sqrt{(-6)^2 + (-4)^2 + (-4)^2 + (-6)^2 + (-6)^2 + (-6)^2} \\ &= \sqrt{36 + 16 + 16 + 36 + 36 + 36} \\ &= \sqrt{176} = 13.27 \end{aligned}$$

Data K7

$$\begin{aligned} D(K_7, C_1) &= \sqrt{(21 - 20)^2 + (20 - 18)^2 + (22 - 22)^2 + (23 - 21)^2 + (24 - 20)^2 + (25 - 19)^2} \\ &= \sqrt{(1)^2 + (2)^2 + (0)^2 + (2)^2 + (4)^2 + (6)^2} \\ &= \sqrt{1 + 4 + 0 + 4 + 16 + 36} \\ &= \sqrt{61} = 7.81 \end{aligned}$$

$$\begin{aligned} D(K_7, C_2) &= \sqrt{(21 - 25)^2 + (20 - 24)^2 + (22 - 23)^2 + (23 - 24)^2 + (24 - 25)^2 + (25 - 26)^2} \\ &= \sqrt{(-4)^2 + (-4)^2 + (-1)^2 + (-1)^2 + (-1)^2 + (-1)^2} \end{aligned}$$

$$\begin{aligned} &(-1)^2 \\ &= \sqrt{16 + 16 + 1 + 1 + 1 + 1} \\ &= \sqrt{36} = 6 \end{aligned}$$

Data K8

$$\begin{aligned} D(K_8, C_1) &= \sqrt{(17 - 20)^2 + (16 - 18)^2 + (18 - 22)^2 + (17 - 21)^2 + (16 - 20)^2 + (15 - 19)^2} \\ &= \sqrt{(-3)^2 + (-2)^2 + (-4)^2 + (-4)^2 + (-4)^2 + (-4)^2} \\ &= \sqrt{9 + 4 + 16 + 16 + 16 + 16} \\ &= \sqrt{77} = 8.77 \end{aligned}$$

$$\begin{aligned} D(K_8, C_2) &= \sqrt{(17 - 25)^2 + (16 - 24)^2 + (18 - 23)^2 + (17 - 24)^2 + (16 - 25)^2 + (15 - 26)^2} \\ &= \sqrt{(-8)^2 + (-8)^2 + (-5)^2 + (-7)^2 + (-9)^2 + (-11)^2} \\ &= \sqrt{64 + 64 + 25 + 49 + 81 + 121} \\ &= \sqrt{404} = 20.10 \end{aligned}$$

Data K9

$$\begin{aligned} D(K_9, C_1) &= \sqrt{(23 - 20)^2 + (22 - 18)^2 + (24 - 22)^2 + (25 - 21)^2 + (24 - 20)^2 + (23 - 19)^2} \\ &= \sqrt{(3)^2 + (4)^2 + (2)^2 + (4)^2 + (4)^2 + (4)^2} \\ &= \sqrt{9 + 4 + 16 + 16 + 16 + 16} \\ &= \sqrt{77} = 8.77 \end{aligned}$$

$$\begin{aligned} D(K_9, C_2) &= \sqrt{(23 - 25)^2 + (22 - 24)^2 + (24 - 23)^2 + (25 - 24)^2 + (24 - 25)^2 + (23 - 26)^2} \\ &= \sqrt{(-2)^2 + (-2)^2 + (1)^2 + (1)^2 + (-1)^2 + (-3)^2} \\ &= \sqrt{4 + 4 + 1 + 1 + 1 + 9} \\ &= \sqrt{20} = 4.47 \end{aligned}$$

Data K10

$$\begin{aligned} D(K_{10}, C_1) &= \sqrt{(20 - 20)^2 + (21 - 18)^2 + (20 - 22)^2 + (21 - 21)^2 + (22 - 20)^2 + (21 - 19)^2} \\ &= \sqrt{(0)^2 + (3)^2 + (-2)^2 + (0)^2 + (2)^2 + (2)^2} \\ &= \sqrt{0 + 9 + 4 + 0 + 4 + 4} \\ &= \sqrt{21} = 4.58 \end{aligned}$$

$$\begin{aligned} D(K_{10}, C_2) &= \sqrt{(20 - 25)^2 + (21 - 24)^2 + (20 - 23)^2 + (21 - 24)^2 + (22 - 25)^2 + (21 - 26)^2} \\ &= \sqrt{(-5)^2 + (-3)^2 + (-3)^2 + (-3)^2 + (-3)^2 + (-5)^2} \\ &= \sqrt{25 + 9 + 9 + 9 + 9 + 25} \\ &= \sqrt{86} = 9.27 \end{aligned}$$

Dari hasil perhitungan diatas mendapatkan hasil bahwa jarak data pertama dengan pusat cluster pertama adalah 0. Berdasarkan hasil kedua

perhitungan diatas bahwa jarak data kedua yang paling dekat adalah dengan cluster 1, sehingga data pertama dimasukkan ke dalam Kluster 1, Sampel Hasil perhitungan selengkapnya untuk 10 data pertama dapat dilihat pada tabel 4.

Tabel 4. Sampel hasil pengelompokkan iterasi pertama

Nama	Januari	Februari	Maret	April	Mei	Juni
K1	20	18	22	21	20	19
K2	22	23	24	25	22	23
K3	18	19	20	21	22	23
K4	16	17	18	17	16	15
K5	25	24	23	24	25	26
K6	19	20	19	18	19	20
K7	21	20	22	23	24	25
K8	17	16	18	17	16	15
K9	23	22	24	25	24	23
K10	20	21	20	21	22	21

Nama	C1	C2	Kluster
K1	0	12.04	C1
K2	8.31	5.48	C2
K3	5.39	10.49	C1
K4	9	20.14	C1
K5	12.04	0	C2
K6	5	13.27	C1
K7	7.81	6	C2
K8	8.77	20.10	C1
K9	8.77	4.47	C2
K10	4.58	9.27	C1

C1: {K1, K3, K4, K6, K8, K10}

C2: {K2, K5, K7, K9}

4. Hitung kembali pusat kluster dengan keanggotaan Kluster yang sekarang. Pusat Kluster adalah rata-rata dari semua data/objek dalam Kluster tertentu

$$\begin{aligned} C1 &= \left(\frac{\sum_{i=1}^n X_i}{n} \right), \left(\frac{\sum_{i=1}^n Y_i}{n} \right), \left(\frac{\sum_{i=1}^n Z_i}{n} \right) \\ &= \left(\frac{\sum_{i=1}^n X_i}{n} \right), \left(\frac{\sum_{i=1}^n Y_i}{n} \right), \left(\frac{\sum_{i=1}^n Z_i}{n} \right) \end{aligned}$$

Tabel 5. Sampel titik pusat awal kluster yang baru

Nama	Januari	Februari	Maret	April	Mei	Juni	C1
K1	18.33	18.5	19.5	19.17	19.1	18.8	C1
K5	22.75	22.25	23.25	24.25	23.7	24.2	C2

5. Selanjutnya melakukan proses

iterasi yang kedua pada data

iterasi pertama

Data K1

$$\begin{aligned} D(K_1, C_1) &= \sqrt{(20 - 18.33)^2 + (18 - 18.5)^2 + (22 - 19.5)^2 + (21 - 19.17)^2 + (20 - 19.17)^2 + (19 - 18.83)^2} \\ &= \sqrt{(1.67)^2 + (-0.5)^2 + (2.5)^2 + (1.83)^2 + (0.83)^2 + (0.17)^2} \\ &= \sqrt{2.79 + 0.25 + 6.25 + 3.35 + 0.69 + 0.03} \\ &= \sqrt{13.36} = 3.65 \end{aligned}$$

$$\begin{aligned} D(K_1, C_2) &= \sqrt{(20 - 22.75)^2 + (18 - 22.25)^2 + (22 - 23.25)^2 + (21 - 24.25)^2 + (20 - 23.75)^2 + (19 - 24.25)^2} \\ &= \sqrt{(-2.75)^2 + (-4.25)^2 + (-1.25)^2 + (-3.25)^2 + (-3.75)^2 + (-5.25)^2} \\ &= \sqrt{7.56 + 18.06 + 1.56 + 10.56 + 14.06 + 27.56} \\ &= \sqrt{79.36} = 8.90 \end{aligned}$$

Data K2

$$\begin{aligned} D(K_1, C_1) &= \sqrt{(22 - 18.33)^2 + (23 - 18.5)^2 + (24 - 19.5)^2 + (25 - 19.17)^2 + (22 - 19.17)^2 + (23 - 18.83)^2} \\ &= \sqrt{(3.67)^2 + (4.5)^2 + (4.5)^2 + (5.83)^2 + (2.83)^2 + (4.17)^2} \\ &= \sqrt{13.49 + 20.25 + 20.25 + 34 + 8.00 + 17.38} \\ &= \sqrt{113.37} = 10.64 \end{aligned}$$

$$\begin{aligned} D(K_1, C_2) &= \sqrt{(22 - 22.75)^2 + (23 - 22.25)^2 + (24 - 23.25)^2 + (25 - 24.25)^2 + (22 - 23.75)^2 + (23 - 24.25)^2} \\ &= \sqrt{(-0.75)^2 + (0.75)^2 + (0.75)^2 + (0.75)^2 + (-1.75)^2 + (-1.25)^2} \\ &= \sqrt{0.56 + 0.56 + 0.56 + 0.56 + 3.06 + 1.56} \\ &= \sqrt{6.88} = 2.62 \end{aligned}$$

Data K3

$$\begin{aligned} D(K_3, C_1) &= \sqrt{(18 - 18.33)^2 + (19 - 18.5)^2 + (20 - 19.5)^2 + (21 - 19.17)^2 + (22 - 19.17)^2 + (23 - 18.83)^2} \\ &= \sqrt{(-0.33)^2 + (0.5)^2 + (0.5)^2 + (1.83)^2 + (2.83)^2 + (4)^2} \\ &= \sqrt{0.11 + 0.25 + 0.25 + 3.35 + 8 + 17.38} \\ &= \sqrt{29.34} = 5.42 \end{aligned}$$

$$\begin{aligned} D(K_3, C_2) &= \sqrt{(18 - 22.75)^2 + (19 - 22.25)^2 + (20 - 23.25)^2 + (21 - 24.25)^2 + (22 - 23.75)^2 + (23 - 24.25)^2} \\ &= \sqrt{(-4.75)^2 + (-3.25)^2 + (-3.25)^2 + (-3.25)^2 + (-1.75)^2 + (-1.25)^2} \\ &= \sqrt{22.56 + 10.56 + 10.56 + 10.56 + 3.06 + 1.56} \end{aligned}$$

$$= \sqrt{58.86} = 7.67$$

Data K4

$$\begin{aligned} D(K_4, C_1) &= \sqrt{(16 - 18.33)^2 + (17 - 18.5)^2 + (18 - 19.5)^2 + (17 - 19.17)^2 + (16 - 19.17)^2 + (15 - 18.83)^2} \\ &= \sqrt{(-2.33)^2 + (-1.5)^2 + (-1.5)^2 + (-2.17)^2 + (-3.17)^2 + (-3.83)^2} \\ &= \sqrt{5.43 + 2.25 + 2.25 + 4.71 + 10.05 + 14.66} \\ &= \sqrt{39.35} = 6.27 \end{aligned}$$

$$\begin{aligned} D(K_4, C_2) &= \sqrt{(16 - 22.75)^2 + (17 - 22.5)^2 + (18 - 23.25)^2 + (17 - 24.25)^2 + (16 - 23.75)^2 + (15 - 24.25)^2} \\ &= \sqrt{(-6.75)^2 + (-5.25)^2 + (-5.25)^2 + (-7.25)^2 + (-7.75)^2 + (-9.25)^2} \\ &= \sqrt{45.56 + 27.56 + 27.56 + 52.56 + 60.06 + 85.56} \\ &= \sqrt{298.86} = 17.30 \end{aligned}$$

Data K5

$$\begin{aligned} D(K_5, C_1) &= \sqrt{(25 - 18.33)^2 + (24 - 18.5)^2 + (23 - 19.5)^2 + (24 - 19.17)^2 + (25 - 19.17)^2 + (26 - 18.83)^2} \\ &= \sqrt{(6.67)^2 + (5.5)^2 + (3.5)^2 + (4.83)^2 + (5.83)^2 + (7.17)^2} \\ &= \sqrt{44.49 + 30.25 + 12.25 + 23.31 + 33.99 + 51.41} \\ &= \sqrt{195.70} = 13.99 \end{aligned}$$

$$\begin{aligned} D(K_5, C_2) &= \sqrt{(25 - 22.75)^2 + (24 - 22.25)^2 + (23 - 23.25)^2 + (24 - 24.25)^2 + (25 - 23.75)^2 + (26 - 24.25)^2} \\ &= \sqrt{(2.25)^2 + (1.75)^2 + (-0.25)^2 + (-0.25)^2 + (1.25)^2 + (1.75)^2} \\ &= \sqrt{5.06 + 3.06 + 0.06 + 0.06 + 1.56 + 3.06} \\ &= \sqrt{12.86} = 3.59 \end{aligned}$$

Data K6

$$\begin{aligned} D(K_6, C_1) &= \sqrt{(19 - 18.33)^2 + (20 - 18.5)^2 + (19 - 19.5)^2 + (18 - 19.17)^2 + (19 - 19.17)^2 + (20 - 18.83)^2} \\ &= \sqrt{(0.67)^2 + (1.5)^2 + (-0.5)^2 + (-1.17)^2 + (-0.17)^2 + (1.17)^2} \\ &= \sqrt{0.45 + 2.25 + 0.25 + 1.37 + 0.03 + 1.37} \\ &= \sqrt{5.72} = 2.39 \end{aligned}$$

$$\begin{aligned} D(K_6, C_2) &= \sqrt{(19 - 22.75)^2 + (20 - 22.25)^2 + (19 - 23.25)^2 + (18 - 24.25)^2 + (19 - 23.75)^2 + (20 - 24.25)^2} \\ &= \sqrt{(-3.75)^2 + (-2.25)^2 + (-4.25)^2 + (-6.25)^2 + (-4.75)^2 + (-4.25)^2} \\ &= \sqrt{14.06 + 5.06 + 18.06 + 39.06 + 22.56 + 18.06} \\ &= \sqrt{116.86} = 10.81 \end{aligned}$$

Data K7

$$\begin{aligned} D(K_7, C_1) &= \sqrt{(21 - 18.33)^2 + (20 - 18.5)^2 + (22 - 19.5)^2 + (23 - 19.17)^2 + (24 - 19.17)^2 + (25 - 18.83)^2} \\ &= \sqrt{(2.67)^2 + (1.5)^2 + (2.5)^2 + (3.83)^2 + (4.83)^2 + (6.17)^2} \\ &= \sqrt{7.12 + 2.25 + 6.25 + 14.67 + 23.31 + 38.06} \\ &= \sqrt{91.66} = 9.57 \end{aligned}$$

$$\begin{aligned} D(K_7, C_2) &= \sqrt{(21 - 22.75)^2 + (20 - 22.25)^2 + (22 - 23.25)^2 + (23 - 24.25)^2 + (24 - 23.75)^2 + (25 - 24.25)^2} \\ &= \sqrt{(-1.75)^2 + (-2.25)^2 + (-1.25)^2 + (-1.25)^2 + (0.25)^2 + (0.75)^2} \\ &= \sqrt{3.06 + 5.06 + 1.56 + 1.56 + 0.06 + 0.56} \\ &= \sqrt{11.86} = 3.44 \end{aligned}$$

Data K8

$$\begin{aligned} D(K_8, C_1) &= \sqrt{(17 - 18.33)^2 + (16 - 18.5)^2 + (18 - 19.5)^2 + (17 - 19.17)^2 + (16 - 19.17)^2 + (15 - 18.83)^2} \\ &= \sqrt{(-1.33)^2 + (-2.5)^2 + (-1.5)^2 + (-2.17)^2 + (-3.17)^2 + (-3.83)^2} \\ &= \sqrt{1.77 + 6.25 + 2.25 + 4.71 + 10.05 + 14.66} \\ &= \sqrt{39.69} = 6.30 \end{aligned}$$

$$\begin{aligned} D(K_8, C_2) &= \sqrt{(17 - 22.75)^2 + (16 - 22.25)^2 + (18 - 23.25)^2 + (17 - 24.25)^2 + (16 - 23.75)^2 + (15 - 24.25)^2} \\ &= \sqrt{(-5.75)^2 + (-6.25)^2 + (-5.25)^2 + (-7.25)^2 + (-7.75)^2 + (-9.25)^2} \\ &= \sqrt{33.06 + 39.06 + 27.56 + 52.56 + 60.06 + 85.56} \\ &= \sqrt{298.86} = 17.30 \end{aligned}$$

Data K9

$$\begin{aligned} D(K_9, C_1) &= \sqrt{(23 - 18.33)^2 + (22 - 18.5)^2 + (24 - 19.5)^2 + (25 - 19.17)^2 + (24 - 19.17)^2 + (23 - 18.83)^2} \\ &= \sqrt{(4.67)^2 + (3.5)^2 + (4.5)^2 + (5.83)^2 + (4.83)^2 + (4.17)^2} \\ &= \sqrt{21.78 + 12.25 + 20.25 + 34.00 + 23.31 + 17.38} \\ &= \sqrt{128.97} = 11.36 \end{aligned}$$

$$\begin{aligned} D(K_9, C_2) &= \sqrt{(23 - 22.75)^2 + (22 - 22.25)^2 + (24 - 23.25)^2 + (25 - 24.25)^2 + (24 - 23.75)^2 + (23 - 24.25)^2} \\ &= \sqrt{(0.25)^2 + (-0.25)^2 + (0.75)^2 + (0.75)^2 + (0.25)^2 + (-1.25)^2} \\ &= \sqrt{0.06 + 0.06 + 0.56 + 0.56 + 0.06 + 1.56} \\ &= \sqrt{2.86} = 1.69 \end{aligned}$$

Data K10

$$D(K_{10}, C_1) = \sqrt{(20 - 18.33)^2 + (21 -$$

$$\begin{aligned} &18.5)^2 + (20 - 19.5)^2 + (21 - 19.17)^2 + (22 - 19.17)^2 + (21 - 18.83)^2} \\ &= \sqrt{(1.67)^2 + (2.5)^2 + (0.5)^2 + (1.83)^2 + (2.83)^2 + (2.17)^2} \\ &= \sqrt{2.79 + 6.25 + 0.25 + 3.35 + 8.00 + 4.71} \\ &= \sqrt{25.35} = 5.03 \end{aligned}$$

$$\begin{aligned} D(K_{10}, C_2) &= \sqrt{(20 - 22.75)^2 + (21 - 22.25)^2 + (20 - 23.25)^2 + (21 - 24.25)^2 + (22 - 23.75)^2 + (21 - 24.25)^2} \\ &= \sqrt{(-2.75)^2 + (-1.25)^2 + (-3.25)^2 + (-3.25)^2 + (-1.75)^2 + (-3.25)^2} \\ &= \sqrt{7.56 + 1.56 + 10.56 + 10.56 + 3.06 + 10.56} \\ &= \sqrt{43.86} = 6.62 \end{aligned}$$

Tabel 6. Sampel hasil uji iterasi ke 2

Nama	Januari	Februari	Maret	April	Mei	Juni
K1	20	18	22	21	20	19
K2	22	23	24	25	22	23
K3	18	19	20	21	22	23
K4	16	17	18	17	16	15
K5	25	24	23	24	25	26
K6	19	20	19	18	19	20
K7	21	20	22	23	24	25
K8	17	16	18	17	16	15
K9	23	22	24	25	24	23
K10	20	21	20	21	22	21

Nama	C1	C2	Kluster
K1	3.65	8.90	C1
K2	10.64	2.62	C2
K3	5.42	7.67	C1
K4	6.27	17.30	C1
K5	13.99	3.59	C2
K6	2.39	10.81	C1
K7	9.57	3.44	C2
K8	6.30	17.30	C1
K9	11.36	1.69	C2
K10	5.03	6.62	C1

C1: {K1, K3, K4, K6, K8, K10}

C2: {K2, K5, K7, K9}

Dalam penelitian ini terjadi 2 kali iterasi pada proses iterasi ke 2 tidak adanya anggota yang berpindah dari kluster satu kekluster yang lain atau pusat kluster tidak berubah, sehingga proses iterasi selesai. Dan hasil dari proses iterasi ke 2 dengan menggunakan atribut jumlah absen perbulan dengan kluster 1 menghasilkan sekelompok data yang berjumlah 6 item dan kluster 2 berjumlah 4 item.

Implementasi Sistem

Pengimplementasian dari data absen siswa SMP Negeri 1 Bilah Barat menggunakan RapidMiner 10.3. RapidMiner merupakan sebuah aplikasi untuk pengolahan data dengan

menggunakan prinsip dan algoritma data mining.



Gambar 2. Rapidminer Studio 10.3

RapidMiner mampu mengekstrak pola-pola dari data set yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan *database* sehingga sangat cocok digunakan untuk metode *K-Means Clustering*. Aplikasi ini dapat menyelesaikan masalah *data mining* diantaranya untuk menghasilkan kelompok – kelompok kluster.

Data dan Teknik Pengujian

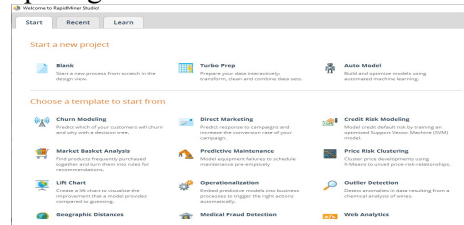
Pada tahap ini dilakukan pengujian data absensi siswa SMP Negeri 1 Bilah Barat yang telah ditransformasi menjadi tabular data. Berikut tabel pengujian data dapat dilihat pada tabel 7.

Tabel 7. Data Pengujian

Nama	Januari	Februari	Maret	April	Mei	Juni
K1	20	18	22	21	20	19
K2	22	23	24	25	22	23
K3	18	19	20	21	22	23
K4	16	17	18	17	16	15
K5	25	24	23	24	25	26
K6	19	20	19	18	19	20
K7	21	20	22	23	24	25
K8	17	16	18	17	16	15
K9	23	22	24	25	24	23
K10	20	21	20	21	22	21

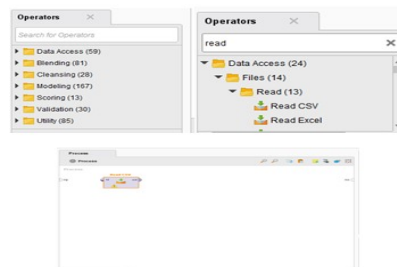
Tabel 7 merupakan data yang akan diolah di *Rapidminer Studio 10.3*. Adapun Langkah awal dalam pengujian yakni dengan menjalankan *Software Rapidminer Studio 10.3*. dilanjutkan dengan mengklik file lalu *new process* dan mengklik sub *menu blank (Start*

with a blank Process). Sehingga muncul seperti gambar 3.



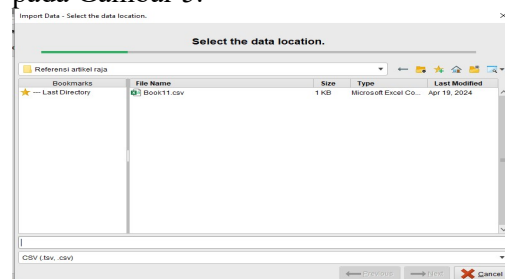
Gambar 3. Langkah Awal Rapidminer Studio 10.3

Selanjutnya melakukan penginputan data dengan mengklik menu operators yang berada pada sudutkiri dan ketik cari *READ CSV dan Tarik ke dashboard process* keseperti Gambar 4.



Gambar 4. Penginputan Data Pada Program

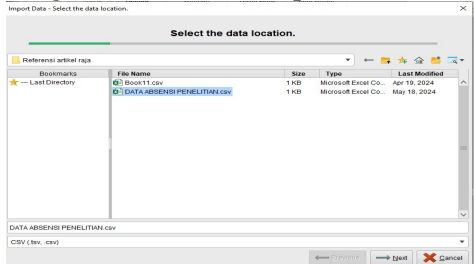
Setelah menarik ke dashboard proses klik Read CSV dua kali maka akan terbuka tampilan lokasi data yang tersimpan, kemudian pilih *database* yang udah ada apabila data yang dimiliki tersimpan di komputer. Seperti pada Gambar 5.



Gambar 5. Pemilihan Lokasi Data

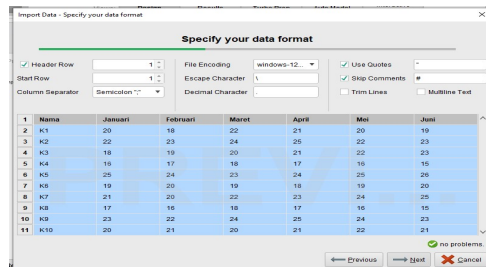
Setelah memilih lokasi data maka langkah selanjutnya mencari

lokasi data yang akan diolah kemudian mengklik tombol *next* seperti Gambar 6.



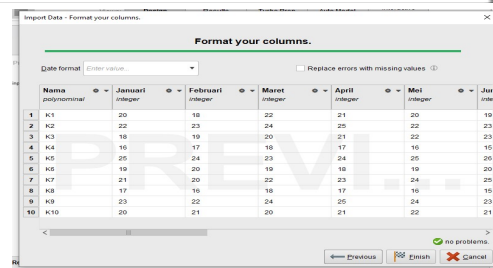
Gambar 6. Pencarian Lokasi Data

Setelah melakukan pencarian data yang akandiinput maka akan terlihat padakolom data yang sudah disiapkan di file *excel* dengan format *.csv* yang sudah berhasil dibaca program *Rapidminer Studio 10.3* Selanjutnya memilih lokasi data yang akan diolah berdasarkan *Sheet* apabila data *excel* yang diperoleh memiliki *Sheet* yang banyak seperti Gambar 8 berikut.



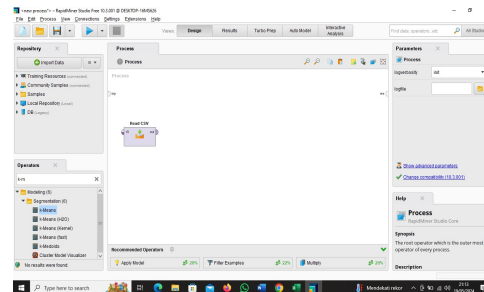
Gambar 8. Pilih Sheet Excel

Setelah memilih lokasi sheet yang akan diolah, langkah selanjutnya memblok *Cell* data yang akan diproses. Kemudian klik *next*, Selanjutnya menunggu data yang dipilih diproses oleh program *RapidMiner Studio*, Setelah data diproses maka program akan menampilkan data yang telah berhasil diinput ke *RapidMiner* seperti Gambar 9 berikut.



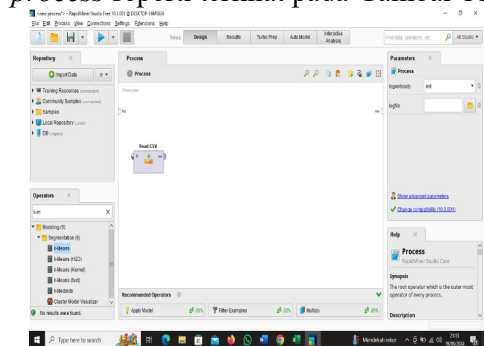
Gambar 9. Preview Data

Selanjutnya setelah data berhasil diinputkan maka, klik *design* pada *menu view* maka akan tampil *form main process* yang merupakan tempat lembar kerja pengolahan data pada *Rapidminer Studio 10.3* untuk lebih jelasnya dapat dilihat pada Gambar 10.



Gambar 10. Form Main Proses

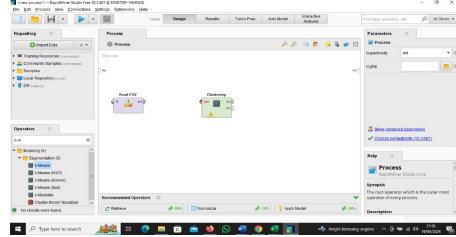
Pada *form main process* pengguna dapat memasukkan data yang akan diproses, sehingga mendapatkan hasil. Pada penelitian ini digunakan *K-Means Clustering* untuk memproses data tersebut. Selanjutnya melakukan *drag* data pengujian kedalam *main process* seperti terlihat pada Gambar 11.



Gambar 11. Tampilan Drag Data Ke

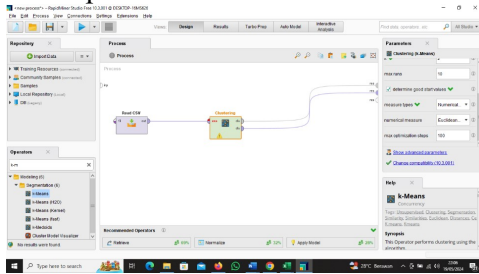
Main Proses

Selanjutnya memasukkan fungsi *K-Means Clustering* dengan cara memilih menu *operators* kemudian pilih *folder K-Means*, *Folder Clustering*, dan mendrag *K-Means Clustering* ke dalam *main process* seperti Gambar 12.



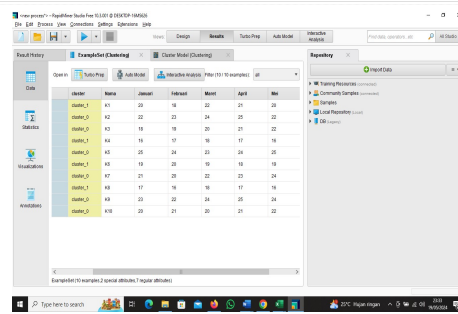
Gambar 12. Mendrag K-Means Clustering Kedalam Main Proses

Kemudian hubungkan *output* dari *Retrieve* data pengujian ke *example-set K-Means Clustering* dan *example-set K-Means Clustering* ke *Result*, kemudian tahapan selanjutnya mengatur parametersnya dimana *K* nya menjadi 2 max runs 10, measure types nya *MixedMeasures*, *Mixed Measures* nya menjadi *Euclidean Distance* selanjutnya jalankan aplikasi dengan mengklik tombol *Run*. Aplikasi akan menampilkan hasil *input* data dan dapat dilihat pada Gambar 13.



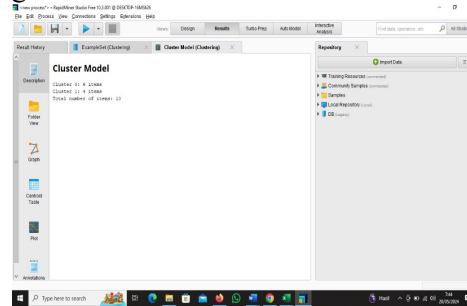
Gambar 13. Hasil Drag K-Means Clustering Pada Main Process Hasil Pengujian

Setelah melewati rangkaian tahapan pengujian pada system sebelumnya, makahasilnya dapat dilihat pada Gambar 13.



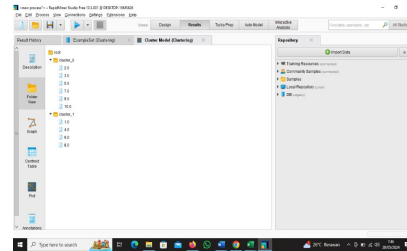
Gambar 13. Example Set

Pada *Example Set* menunjukkan hasil dari kombinasi item set dengan menentukan kelompok cluster pada data tersebut. Tahapan selanjutnya yaitu melihat hasil dari cluster modelnya dengan mengeklik menu *Cluster Model* pilih *Description* dan dapat dilihat pada gambar 14.



Gambar 14. Cluster Model Data Absensi SMP Negeri 1 Bilah Barat

Selanjutnya melihat hasil dari *Folder View Cluster model* pada Gambar 15.



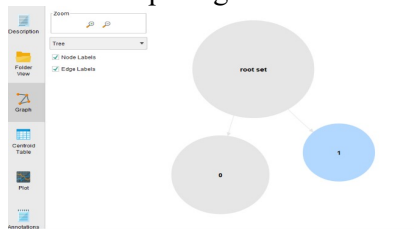
Gambar 16. Foder View Cluster Model

Selanjutnya memperlihatkan hasil dari *Centroid Tabelnya* pada *Cluster Model* dilihat pada gambar 17.

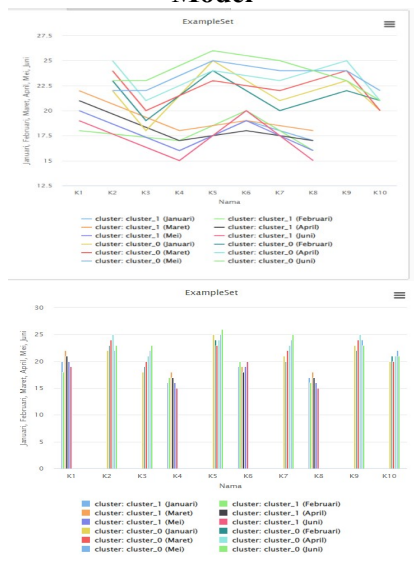
	cluster_0	cluster_1
nama	4760	4760
amen	2100	9
nama	2100	9760
amen	2100	9760
amen	2100	9760
amen	2100	9760
amen	2100	9760
amen	2100	9760
amen	2100	9760

Gambar 17. Centroid Table Cluster Model

Selanjutnya memperlihatkan hasil akhir dari diagram plotnya pada Cluster Model pada gambar 18.



Gambar 18. Diagram Plot Cluster Model



Gambar 19. Diagram Plot Example Set

Berdasarkan pengujian menggunakan aplikasi *Rapidminer* maka dapat disimpulkan bahwa pemilihan jumlah kluster sebanyak 2. Maka hasil pengujian menunjukkan

bahwa kluster 0 adalah kluster 1 berjumlah 6 item, kluster 1 adalah kluster 2 berjumlah 4 item. Hasil Clustering data absensi siswa SMP Negeri 1 Bilah Barat dapat mengungkapkan pola absensi yang bermanfaat bagi pendidik dan sekolah. Analisis terhadap *cluster – cluster* ini dapat mengungkapkan informasi tentang absensi siswa, kecenderungan dalam keberhasilan atau kesulitan dalam sekolah dan dalam mata pelajaran, serta memungkinkan identifikasi siswa mana yang sering Sekolah dan malas sekolah yang memerlukan bantuan tambahan.

V. Kesimpulan

Tujuan analisis data mining menggunakan algoritma K-means Clustering dalam evaluasi absensi siswa Smp negeri 1 bilah barat adalah untuk mengidentifikasi pola – pola absensi siswa dalam sekolah, membagi siswa ke dalam kelompok – kelompok berdasarkan tingkat kehadirannya dan tingkat kerajinan dalam sekolah, serta memberikan wawasan yang berharga kepada guru dan sekolah. Hasil Clustering data absensi siswa Smp negeri 1 bilah barat dapat mengungkap pola yang bermanfaat bagi pendidik dan administrator sekolah. Analisis terhadap cluster – cluster ini dapat mengungkapkan informasi tentang kehadiran siswa, kecenderungan keberhasilan atau kesulitan dalam mutu belajar sekolah, serta memungkinkan identifikasi siswa yang rajin dan tidak rajin. Pengelompokan hasil cluster berdasarkan data absensi siswa dengan K-means memperoleh 2 kelompok siswa yaitu siswa Rajin dengan kelompok cluster 0 dan kelompok siswa yang tidak rajin dengan kelompok cluster 1. Kelompok C0 siswa rajin

terdiri dari 6 siswa dan kelompok C1 terdiri dari 4 siswa.

VI. Daftar Pustaka

- Anggada Maulana, "Konsep Dasar Data Mining," Konsep Data Mining, vol. 1, pp. 1–16, 2018.
- Arsyad, "Bab II kajian teori," BAB 2 Kajian teori, no. 1, pp. 16–72, 2017.
- B. A. B. II, "Tahapan Knowledge Discovery in Databases (KDD)," pp. 1–11, 2007.
- D. Hidayatullah, "Chemical Information and Modeling," Journal of Chemical Information and Modeling, vol. 53, no. 9, pp. 8–24, 2018.
- D. M. Budanis and A. Wisnu Wardana, "Sistem Pendukung Keputusan Penentuan Resep," Jurnal Ilmiah NERO, vol. 5, no. 1, pp. 47–57, 2020.
- D. Tilang and D. I. Instansi, "Media Informatika Vol.20 No.2 (2021) 109," vol. 20, no. 2, pp. 109–121, 2021.
- F. J. Martínez-López and J. Casillas, "KDD /Data Mining ," Wiley Encyclopedia of Management, pp. 1–1, 2015, doi: 10.1002/9781118785317.weom090075.
- K. Handoko, "Penerapan Data Mining Dalam Meningkatkan Mutu Pembelajaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus Di Program Studi Tkj Akademi Komunitas Solok Selatan)," Jurnal Teknologi dan Sistem Informasi, vol. 02, no. 03, pp. 31–40, 2016, [Online]. Available: <http://teknosi.fti.unand.id/index.php/teknosi/article/view/70>
- M. Benri, H. Metisen, and S. Latipa, "Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila," Jurnal Media Infotama, vol. 11, no. 2, pp. 110–118, 2015, <https://core.ac.uk/download/pdf/287160954.pdf>
- N. P. Wulandari, "Studi Identifikasi Faktor-Faktor yang Mempengaruhi Stres Belajar Siswa Full Days SMP Muhammadiyah 01 Medan," Universitas Medan Area, pp. 12–26, 2018.
- T. Alfina, B. Santosa, and A. R. Barakbah, "Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data," Teknik Its, vol. 1, no. 1, pp. 521–525, 2012, <http://www.ejurnal.its.ac.id/index.php/teknik/article/view/1794>
- W. Lestari, "Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus: STMIK Bina Bangsa Kendari)," Simkom, vol. 4, no. 2, pp. 35–48, 2019, doi: 10.51717/simkom.v4i2.37
- W. SETIA, "... Dengan Pendekatan Client Centered Therapy Dalam Mengatasi Kesulitan Belajar Peserta Didik Kelas Viii Di Smp Negeri 2 Batu ...," 2022.